



A:Accession: A81931  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-556 <PAR>  
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84134.1; PID:g737956  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: mafB3; NMA0853

alignment\_scores:  
Quality: 2200.00 Length: 560  
Ratio: 4.593 Gaps: 8  
Percent Similarity: 85.536 Percent Identity: 77.500  
alignment\_block:  
US-09-303-518D-463 x A81931

Align seg 1/1 to: A81931 from: 1 to: 556

1 TTGGGCAATTCCTCCGCAAAATATCCCTTATTCGTCCATCTGTCAGTGTG 50  
:::|||||  
1 MetGlyLeuSerArgLysIleSerLeuIleLeuSerIleLeuAlaValCy 17  
51 CTGCGCATGTCATGCACGCTCAGATTGGCAACCATCCCTTTATCC 100  
|||||  
17 sLeuProMetHisAlaHisAlaSerAspLeuAlaAsnAspSerPheIleA 34  
|||||  
101 GGCAGGTTCTCGACCTCAGCATTTTCGAACCCGAGCGGAAATACACCTA 150  
34 rGlnValLeuAspArgGlnHisPheGluProAspGlyLysTyrHisLeu 50  
151 TTCCGACGAGGGGAGGTGTCGACGACGACGACGACGACGACGACGACG 200  
51 PheGlySerArgGlyGluLeuAlaGluArgSerGlyHisIleGlyLeuG 67  
201 AACATCAAGCATCAGTTGGGCGACCTGATGATTCACAGCGCGCG 250  
67 yAsnIleGlnSerHisGlnLeuGlyAsnLeuPheIleGlnAlaAlaI 84  
251 TTGAAGGAATATCGCTACATGTCGCTTTCGATCCGATCCAGGACAAA 300  
84 LeGlyAsnIleGlyTyrIleValArgPheSerAspHisGlyHisGlu 100  
301 TTCCATTCGCGCTTCGACACCATGCTCATCTCCGATTCGACGAGC 350  
101 ValHisSerProPheAspAsnHisAlaSerHisSerAspSerAspGluAl 117  
351 CGGTAGTCCGCTTCGACGATTCAGCCTTTTACCCCATCCATTTGGGCGGAT 400  
117 aGlySerProValAspGlyPheSerLeuTyrArgIleHisTrpAspGlyT 134  
401 ACGAACCATTCCTCCGCGGCTATGACGGCGCACAGGCGCGGCTAT 450  
134 yrGluHisHisProAlaAspGlyTyrAspGlyProGlnGlyGlyTyr 150  
451 CCCGCTCCCAAGCGCGGAGGATATACAGTACGATACGACATAAAGCGCT 500  
151 ProAlaProLysGlyAlaArgAspIleTyrSerTyrAspIleLysGlyVa 167  
501 TGCCCAAAATATCCGCTCAACCTGACGACGACGACGACGACGACGACG 550  
167 lAlaGlnAsnIleArgLeuAsnLeuThrAspAsnArgSerThrGlyGlnA 184  
551 GCCTTGGCGCGCTTTCACAAATCCGCGCTATGCTGACGACGAGGAGTA 600  
184 rGLeuValAspArgPheHisAsnThrGlySerMetLeuThrGlnGlyVal 200  
601 GCGCAGGATTCAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATC 650  
201 GlyAspGlyPheLysArgAlaThrArgTyrSerProGluLeuAspArgSe 217  
651 GGGCAATCCCGCCGAGCCTTCACGCGCAGCTGTCAGATATCGTCAAAAACA 700

217 rGlyAsnAlaAlaGluAlaPheAsnGlyThrAlaAspIleValLysAsnI 234  
701 TCATCGCGCGCGGAGGAGAAATTCCTCGCGCAGCGCATCGCTGCAGGGT 750  
234 leIleGlyAlaAlaGlyGluIleValGlyAlaGlyAspAlaValGlnGly 250  
751 ATAAGCGAAGCTCAACATTCGTGTCTCATCGCGCTTGGCTTGGCTTTC 800  
251 IleSerGluGlySerAsnIleAlaValMetHisGlyLeuGlyLeuLeuSe 267  
801 CACCGAAACAAAGATGGCGGCATCAACGATTTGGCAGATATGGCGCAAC 850  
267 rThrGluAsnLysMetAlaArgIleAsnAspLeuAlaAspMetAlaGlnL 284  
851 TCAAAGACTATGCGCGCAGCAGCATCCGCGATTGGCGAGTCCAAACCCC 900  
284 euLysAspTyrAlaAlaAlaIleArgAspTrpAlaValGlnAsnPro 300  
901 AATCGCGCACAGGCATAGAACCGCTCAGCAATATCTTTATGGCAGCAT 950  
301 AsnAlaAlaGlnGlyIleGluAlaValSerAsnIlePheThrAlaValI 317  
951 CCCCATCAAGGATTTGAGCTGTCGCGGGAATAACGGCTTGGCGGCA 1000  
317 eProValLysGlyIleGlyAlaValArgGlyLysTyrGlyLeuGlyI 334  
1001 TCACGCGCATCTCTCAAGCGCTCGCAGATGGCGGCGCATTCGCTCCG 1050  
334 lThrAlaHisProValLysArgSerGlnMetGlyGluIleAlaLeuPro 350  
1051 AAAGGAAATCCCGCTCAGCAGCAATTTTCCGATGGCGCATACGCCAA 1100  
351 LysGlyLysSerAlaValSerAspAsnPheAlaAspAlaIleTyrAlaLy 367  
1101 ATACCGCTCCCTTACCATTCCGAAATATCGTTCAAACTTGGAGCAGC 1150  
367 sTyProSerProTyrHisSerArgAsnIleArgSerAsnLeuGluInA 384  
1151 GTTACGGCAAGAAACATCCTCTCAACCGCTCCGCGCGCTCAACCGC 1200  
384 rGtyrGlyLysGluAsnIleThrSerSerThrValProProSerAsnGly 400  
1201 AAAATGTCAACTGCACCAACCCACCCAGAGACAGCGGTACCGTT 1250  
401 LysAsnValLysLeuAlaAsnLysArgHisProLysThrLysValProPh 417  
1251 TCACGCTAAAGGGTTTCCGAAATTTGAGAAGCAGCGTGAATATGATACA 1300  
417 eAspGlyLysGlyPheProAsnPheGluLysAspValLysTyrAspThra 434  
1301 AGCTCGAT.....ATCAAGAATTATCGGGGCGGTATACCTAAGGCT 1344  
434 rGleAsnThrAlaValProGlnValAsn.....ProIleAsp 446  
1345 AAGCTCTGTTTGTATGCGAAA.....CCGAGATGGGA 1376  
447 GluProValPheAsnProLysGlySerValGlySerAlaHisSerTrpSe 463  
1377 GGTGTATAGGAAGCTTAAT.....AAATTGACAACTCGTGCAGCGTGG 1420  
463 rIleThrAlaArgIleGlnTyrAlaLysLeuProArgGlnGlyArgIleA 480  
1421 AG.....AAAAATGTCAGGAAACAGAGA..... 1443  
480 rGtyrIleProProLysAsnTyrSerProSerAlaProLeuProLysGly 496  
1443 ..... 1443  
497 ProAsnAsnGlyTyrLeuAspLysPheGlyAsnGluTrpThrLysGlyPr 513  
1444 ..AGAGGAGTCAGACTAGTCAGTTTAAGCCCATGCGCAA..... 1482  
|||.....:|||||



140 lshisProAlaAspAlaTyrAspGlyProLysGlyAsnTyrProLys 156  
457 CCCAAGCGCGGGGATATACAGTACGACATAAAAGCGGTGCCCA 506  
157 ProThrGlyAlaA-gaspGluTyrThrTyrHisValasnGlyThrAlaAr 173  
507 AAATATCCGCTCAACCTGACCGCAACCGCGACCGGACACAGCGGTG 556  
173 gserileLysLeuasnProThrAspThrArgSerIleArgGlnArgile 190  
557 CGGACCGGTTTCCAAATCCCGCGCTATGTCAGCGAAGGAGTAGGCGAC 606  
190 erAspAsnTyrSerAsnLeuGlySerAsnPhSerAspArgAlaAspGlu 206  
607 GGNATCAACAGCGGCCACCGATAGACGCCCGAGCTGGACAGATCGGCAA 656  
207 AlaAsnArgLysMetPheGluHisasnAlaLysLeuAspArgTrpGlyAs 223  
657 TGCCGCGGAGGCTTCAACGGCAGTGCAGATATCGTCAAAACATCATCG 706  
223 nSerMetGluPheIleAsnGlyValAlaAlaGlyAlaLeuAsnPropHeI 240  
707 GCGCGCGAGAGAAATTTGCGCGCGAGCGGATCGCGTCAGGGTATAAGC 756  
240 leSerAlaGlyGluAlaLeuGlyIleGlyAspIleLeuTyrGlyThrArg 256  
757 GAAGGCTCAACATTTGCTGTCATGACCGCTTGGGTCTGCTTCCACCGA 806  
257 TyrAlaIleAspLysAlaAlaMetArgAsnIleAlaProLeuProAlaGl 273  
807 AAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAG 856  
273 uGlyLysPheAlaValIleGlyLeuGlySerValAlaGlyPheGluL 290  
857 ACTATGCGCGAGCAGCATCCGCGATTTGGGCGAGTCCAAACCCCAATGCC 906  
290 ysAsnThrArgGluAlaValAspArgTrpIleGlnGluAsnProAsnAla 306  
907 GCACAAGGATGAAGCGCGTACGCAATATCTTTATGTCGCGAGCATCCCAT 956  
307 AlaGluThrValGluAlaLeuValAsnValLeuProPheAla..... 320  
957 CAAGGATGGAGCTGTCGCGGGAATAACGCTTGGGCGGCATCACGG 1006  
320 ..... 320  
1007 CACATCTGTCAAGCGGTGCGAGATGGCGCGATCGCATTCGCGAAGGG 1056  
321 .....LysValLysAsnLeuThrLysAlaAlaLysPro..Gly 332  
1057 AAATCCCGCGTACGCGACATTTTGGCGGATGCGGCATACGCCAAATACCC 1106  
333 LysAlaAlaValSerGlyAspPheSerAlaAla..... 343  
1107 GTCCCTTACCATTCGCGAATATCGGTTCAAACTTTGGAGCAGCGTTAGC 1156  
344 .....TyrAsnThrArgThrArgLysValThrThrGluGlu 358  
1157 GCAAGAAACATCATCTCTCAACCGTCCGCGCTCAACAGCGCAAAAT 1206  
358 llyLeuAsnArgIleArg.....GlnAsnGlnLysAsn 368  
1207 GTCAACTGGCAGACCAACGCCGCCGAGACAGCGGTACCGTTTGACGG 1256  
369 SerAsnIleHisGluLysAsnTyr.....G1 377  
1257 TAAAGGTTTCCGAATTTTGAGAAGCAGGTGAATATGATACGAGCTCG 1306  
377 yargAspAsnProAsn.....HisIleAsnValLeuSerGlyAsn 391  
1307 ATATTCAGAATTA.....TCGGGGCGGTATACCTAAG 1341  
391 erIleGlnHisIleLeuTyrGlyAspGluAlaGlyGly..... 404

381 etLeuLysAspAsnLysTrpGlnGlyThrSerLysSerGlyIleLysile 397  
1840 GAAGGATTTACCAACCTTAATAAGACACATATCCATTTATGAA 1884  
398 GluGlyPheThrGluProAsnArgThrAlaTyrProIleTyrGlu 412  
seq\_name: p1r2:H81782  
seq\_documentation\_block:  
adhesin MafB2 NMA2113 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: H81782  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:2022556  
A:Accession: H81782  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-498 <PAR>  
A:Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85327.1; PID:g738073  
C:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: mafB2; NMA2113  
alignment\_scores:  
Quality: 740.50 Length: 586  
Ratio: 2.165 Gaps: 18  
Percent Similarity: 58.362 Percent Identity: 32.253  
alignment\_block:  
US-09-303-518D-463 x H81782 ..  
Align seg 1/1 to: H81782 from: 1 to: 498  
28 ATTCTGCTCCATCTGGCAGTGTGC.....CTGCCGAT 59  
..... 111  
7 LeuileLysLeuLeuAlaAlaCysAlaValAlaAlaAlaLeuIleGI 23  
.....  
60 GCATGACACGCTCAGATTGGCAACGATCCCTTTATCCGCGAGTTC 109  
.....  
23 nProAlaLeuAlaAspLeuAlaGlnAspProPheIleThrAspAsnA 40  
.....  
110 TCGACCGCTCAGCATTCGAACCGCGAGGAAATACACCTATTTCGCAGC 159  
.....  
40 laGlnArgLysHisThrGluProGlyGlyLysTyrHisLeuPheGlyAsp 56  
.....  
160 ...AGGGGGAGCTTGCNAGCGCAACGCGCATCGATTGGGAACAT 206  
.....  
57 ProArgGlySerValSerAspArgThrGlyGlnIleAsnValIleGlnAs 73  
.....  
207 ACAAGACCATCAGTTGGCCACCTCATGATCAACAGCGCGCGGTGAAG 256  
.....  
73 pThrHisArgMetGlyAsnLeuLeuIleGlnAlaAsnIleAsnG 90  
.....  
257 GAATATCGGCTACATTCGCGCTTTCCGATCAGCGGCGACAAATCCAT 306  
.....  
90 lYThrIleGlyThrHisThrArgPheSerGlyHisGlyTyrGluGluHis 106  
.....  
307 TCGCCCTTCGCAACCATCGCTTCATTCGATTCCTGACGAAGCCGGTAG 356  
.....  
107 AlaProPheAspAsnHisAlaAlaAspSerAlaSerGluGluLysGlyAs 123  
.....  
357 TCCCGTTGACGATTCAGCTTTACCGATCCATTCGACGAGGATACCAAC 406  
.....  
123 nValAspGluGlyPheThrValTyrArgLeuAsnTrpGluGlyHisGluH 140  
.....  
407 ACCATCCCGCGCGCTATACGCGGCACAGCGCGGTATCCCGCT 456  
.....













```

92 rValProProThr...ThrThrSerThrThrThrThrThrValPro 108
330 ACATT.....CCGATTCTCAGCAAGCGGTAGTCGCGTTGACGAGT 370
108 roThrThrThrSerThrThrThrThrThrValProProThrThrThr 124
371 TCAGCCCTTACCGCATCCATT.....GGACGGA 399
125 SerSerThrThrThrThrThrValProProThrThrThrSerSerThrTh 141
400 TACGAACACCATCCGCGCGAGCTATGACGGCCACAGGCGCGCGGCTA 449
141 rThrThrThrValProProThrThrThrThrSerThrThrThrThrTh 158
450 TCCCGCTCCCAAGGCGCGAGGATATACAGCTACGACATAAAGCGC 499
158 alPro...ProThrThrThrSerThrThrThrThrThrThrThrThr 170
500 TTGCCCAAAATATCCGCCTCAACCTGACGCAACCGCAGCACCAGCAAA 549
171 ValProAlaThrThrThrSerThrAlaThrThrThrThrValPro.... 185
550 CGGCTTCCCGCAGCGTTCCACAATGCCGGCGTATGCTGACGCAAGGAGT 599
186 .....ProThrThrSerThr..... 190
600 AGCGCAGCGGATTCAAACGGCGCCACCCGATACAGCCCGCAGCTGGACAGT 649
191 ..ThrThrThrThrThrValProProThrThrThrSerThrThrThrThr 206
650 CGGCAATCGCGCGAAGCCTTCAACGGCACTGCAGATATCCTCAAAAC 699
207 ThrThrValProProThrThrThrSerSerThrThrThrThrThrThr 223
700 ATCA 703
223 rThr 224

```

seq\_name: pir2:T45134

seq\_documentation block:

seq\_documentation\_block:  
hypothetical protein [imported] - Microbacterium ammoniaphilum (fragment)

C;Species: Microbacterium ammoniaphilum

C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000

C;Accession: T45134

R; Striebel, H.M.; Seeber, S.; Jarsch, M.; Kessler, C.

Gene 172, 41-46, 1996

**A;Title: Cloning and characterization of**

A; Reference number:

A;Accession: T45134

A;Status: preliminary

A; Molecule type: DNA

A; Residues: 1-529 <STR>

A; Cross-references: EMBL: X79027: 1

alignment scores:

```
alignment_scores:
  quality: 161.50
  length: 407
```

Quality: Ratio:

Ratio:	1.009	Gaps:	19
Percent Similarity:	39.312	Percent Identity:	34.070

alignment block:

alignment\_block:  
US-09-303-518D-463 x T4513A

Align seg 1/1 to: T45134 from: 1 to: 529

81 GGCAAGCGATCCCTTTATCCGCCACCTTC

81 GGCAACGATCCCTTTATCCGGCAGTTCT.....CGACC 11

[illegible]

58 Gly<sup>1</sup>ThrGly<sup>2</sup>ThrIleGluProArgGlySerHisTyrArgHisArgArgP

74 oala.....GlnArgArgGlnValHisProLeuGlnArgProHis 88  
157 AGCAGGGGGGA..... 167  
|||:| |||  
88 lngluProGlyAlaLaargGlyGluLeuProValArgAspAspArgAlaGlu 104  
168 .....GTTGCCNAGCGCAACGGCC 187  
||| |||:| |||  
105 ArgGlyArgArgGluProAlaArgProAlaAlaArgAspAlaArgArgHi 121  
188 ATATCCGATTGGGAACATACAAACCATCAGTTGGCCACCTTGATGATT 237  
|||:| |||:| |||  
121 sLeuArgLeuGlyAlaHis.....ProA 129  
238 CAACAGCGCGCGTGAAGAAATATCGGTACATTGCTCCGCTTTCCGA 287  
:| |||:| |||:| |||  
129 laargArgGlyValLeuArgArgHisArgArgHis.....Arg 141  
288 TCACGGGCACAAATTCCATTGCGCCTTGCACAAACCATGCCTCACATCCG 337  
:| |||:| |||:| |||  
142 AlaArgGlyglu..... 145  
338 ATTCTGACGAAGCGCGTAGTCCCGTTGACGGATTACGCCTTTACCGCATC 387  
|||:| |||:| |||:| |||  
146 .....ArgGlyArgGlyProArgGlnGlnVal...ProArgGlnHisP 159  
388 CATTGGGACGGATACGAACACCATCCCGCCGACGGCTATCAGCGGCACA 437  
|| ||||| |||:| |||  
159 roArgGlyArgArgAspArgAlaGlyArgProGlyLeu..... 171  
438 GGCGCGCGCTATCCGCTCCCAAGCGCGGAGGATATACAGTACG 487  
||||| |||:| |||:| |||:| |||:| |||  
172 HisArgArgArgAlaAlaArgArgGlnGlyArgProGlnValAr 188  
488 ACATAAAGCGGCTTGCCCAAAATATCCGCTCAACCTGACCGACAACCCG 537  
||||| |||:| |||:| |||  
188 gHis.....GlyAspSp 193  
538 AGCAGCGGACAACGGCTTCCGACCGTTCACAAATCCGCGCTATGCT 587  
|||||:| |||:| |||:| |||  
193 lnhisAlaAspProArgArgProArgspProArgAlaGlyHisHis 209  
588 GACGCAAGGATGAGCGACGGATTCAACGGCGCACCGGATACAGCCCG 637  
|||:| |||:| |||:| |||:| |||:| |||  
210 ProLeuArgGluGlyGlyGlnGlyGlnGluAlaArgProLeuArgAlaAr 226  
638 AGCTGGACAGATCGGS..... 653  
|||:| |||:| |||:| |||  
226 gGlyGlyGlnGlyGlyProGlyProAlaGlyArgHisProAlaLeuG 243  
654 .....CAATGCCGCGCAGCGCTTCAACGG 677  
||| |||:| |||:| |||  
243 lyAspGluAspProArgAlaArgGlnGlyAlaArgAlaAlaHisArg 259  
678 CACTGCAGATATGTCAAAAACATCATCGCGCGGCGAGGAGAAAT...TG 724  
:| |||:| |||:| |||:| |||:| |||:| |||  
260 GluAlaLeuHisLeuArgLeuGlnArgArgArgGlyArgAlaAspGlyCy 276  
725 TCGG.....CCAGGCGCATGCCGTGACGGGTATACCGAA 759  
||||| |||:| |||:| |||:| |||  
276 sarGAlaGlnGlyArgAlaArgArgAlaArgAlaGly..... 289  
760 GGCTCAACATTGCTGTCTATCAGCGCTGGGTCTGCTTCCACCGAAAA 809  
|||:| |||:| |||:| |||:| |||  
290 .....GlyGlyGlyLeuProArgArg 296  
810 CAAGAT.....GGCGCGCATCAA..... 827  
:| |||:| |||:| |||:| |||  
297 GluAspGValArgAlaHisArgProArgProArgGlyArgGlyGlyAl 313  
828 .....CGATTGGCAGATATCGCGCA...ACTCAAGACTAT 861  
||| ||| ||| ||| ||| |||  
313 aAlaArgLeuAspArgAlaGlyGlyValGlyAlaArqProAlaArqProA 330





A>Note: the gene encoding this protein overlaps uvra gene  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

```

alignment_scores:
  Quality: 150.00      Length: 550
  Ratio: 0.708        Gaps: 33
  Percent Similarity: 38.545  Percent Identity: 24.364

alignment_block:
US-09-303-518D-463 x JQ0405
Align seg 1/1 to: JQ0405 from: 1 to: 1106

51 CCTGCCGTCATGCACAGCCCTCAGATTGGCAACAGCATCTT.....95
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 ProProGlyHisThrAspArgLeuArgAlaGlyArgArgProLeuArgPr 617
   95 .....
617 oArgArgAlaValHisArgProAlaProAlaGlyGlnProProHisA 634
   96 .....TATCCGCGAGGTTCGCA.....CCGTCACATTTCGA 128
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
634 rgAspProProAlaProAlaGlyProArgGlnHisProHisArgArg 650
   129 ACCCGACGGGAATACCACTATTTCGGCAGCAGGGGGAGCTTGCNAGC 178
   :||| ||||| :||| :||| :||| :||| :||| :||| :|||
651 AlaArgGlyHisAsp.....ArgArgGlyGlyLeuAspArgGl 664
   179 GCAACGGCCATATCGGATTGGGAACATCAAAAG.....CCATCAGTTG 222
   :||| ||||| ||| ||||| :||| :||| :||| :||| :|||
664 yHisArgProSerArg...GlyArgValArgArgArgGlyArgAlaLeuG 680
   223 GGCACCTGATGATTCAACAGCGGC.....CGTTGAAGGAATAT 263
   || |||| |||| |||| |||| |||| |||| |||| ||||
680 lLeuProGlyGlySerGlnGlyGluHisAlaValArgHisArgArgLeu 696
   264 CGCTACATTGTCGGCTTTTCGCA..... 287
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
697 ProLeuArgProProLeuHisArgGlyAlaGlyAlaAlaSerArgProGl 713
   288 .....TCACGGGCACAATTCCTATCCCTTCGCAACACCATCGCTCA 330
   :||| ||||| :||| ||||| ||||| ||||| ||||| |||||
713 yGluGlyAlaArgAlaAspGlyPro.....ArgArgPro.GlyGlu 726
   331 CATTCGGATTCTGACGAAGCCGTAGTCCGTTGACGGA..... 369
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
727 GlnProGlyArgLeuGlyProGlyProAlaArgGlyProHisGlyAr 743
   370 .....TTCAGCTTTACCGCATCCATTGGGACGGATACGAACACC 409
   ||| :||| :||| :||| :||| :||| :||| :||| :|||
743 gAspGlyArgValArgLeuArgGlnValHis..... 753
   410 ATCCCGCCGACGGCTATGACGGGCCACAGGGCGGGCTATCCGCTCC 459
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
754 ....AlaAspGlnArgAspProLeuGlnGlyProGlyGlnProAla..G 768
   460 AAAGGCGCGAGGATATATACAG.....CTAGCACATAAAGCGCT 500
   ||||| ||||| :||| :||| :||| :||| :||| :||| :|||
768 lNArgArgGlnAlaArgAlaArgProProAlaProValArgGlyThrArg 784
   501 TGCCCA.....AAATATCCGCTCAACCTGACCGA..... 530
   :||| :||| ||||| ||||| ||||| ||||| ||||| |||||
785 AlaProGlyGlnGlyProArgGlyProGluProHisArgAlaHisAl 801
   531 .....CAACCGCACCGGACAAACGGCTTGGCCGACCGTTTCCACAT 573
   ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
801 aThrLeuGlnProArgHisLeuHisGlyArgValArgAspProGlnA 818
   574 GC..... 575
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
818 lAlaLeuArgGlyAspProArgGlyGlyGlyProGlyLeuProAlaGlyPro 834

```

```

576 .....ValLeuLeuGlnHisGlnGlyArgAlaLeuArgGlyValArgArg 851
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
581 CTATGCTGACCAAGGACTAGG.....CGACGGATTC 612
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
851 gHisAlaGluAspArgGluLeuProAlaGlyArgLeuArgAlaValA 868
   613 AA.....ACCGCCACCCGATACAGCC.....CGAGCTGGACAG 647
   :||| ||||| :||| :||| :||| :||| :||| :||| :|||
868 rgGlyValProArgGlyProValGlnProGlyAspAlaArgGlyHisLeu 884
   648 ATCGGCAATCGCCGGAAG..... 668
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
885 GlnGlyGlnGluHisArgArgGlyProArgHisAlaAspArgGlyGlyAr 901
   669 .....CTTCAACGG.....CACTGCA 684
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
901 gGlyLeuLeuGlnArgValHisProHisLeuAlaValProGlyHisAlaA 918
   685 GATATCGTCAA.....AAACATCATCGG 707
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
918 rgArgArgArgSerGlyLeuArgProSerGlyProAlaArgHisHisAla 934
   708 CGCGCAGGAGAAATTCGCGCGCAGCGATGCGTGCAGGGTATAAG.. 755
   ||||| ||||| :||| :||| :||| :||| :||| :||| :|||
935 LeuGlyArgArgGlyProAlaArg.....GluAlaGlyGlyArgAl 948
   756 ....CGAAGGCTCAACATTCTCTCATCGACGGCTTGGTCTGCTTCC 801
   :||| ||||| :||| ||||| ||||| ||||| ||||| |||||
948 aProGluAlaLeuGlnArgProHisHisLeuArgAlaGlyArgAlaAspH 965
   802 AC..... 803
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
965 isGlyValAlaLeuArgArgHisProGlnAlaProAlaArgAlaSerVal 981
   804 .....CGAAACAAGATGCGCGCATCAACGATTTCGCAGATATGG 844
   :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
982 ProArgGlyGlnGlyGlnHisGlyAlaHisHisArg..... 993
   845 CGCAACTCAAGACTATCCCGCAGCAGCCATCCCGATTGGCGAGTCCAA 894
   ||||| ||||| :||| :||| :||| :||| :||| :||| :|||
994 .AlaGlnProArgArgAspGlnGluArgGlyProArg...AspArgProA 1009
   895 AACCC.....CAATGCGCACAAGCAT..... 917
   :||| :||| :||| ||||| ||||| ||||| ||||| |||||
1009 rgProGlyGlyArgLeuArgArgArgHisAspArgGlyHisGlyHisAla 1025
   918 .....AGAAGCGTCAGCAATATCTTTATGGCAGCCATCCCATCA 958
   ||||| :||| :||| :||| :||| :||| :||| :||| |||||
1026 GlyGlyGlyArgAlaArgArgGluProHisGly..ProValProArg 1041
   959 AAGGATTGGAGCTCCGGGGAATAATACGGCTTGGCGGCATCACGGCA 1008
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1042 .....GlyAlaAlaArgValAlaTrpProProThrGl 1053
   1009 CATCTGTCAAGCGTCGAGATGGCGCATTCGATTGCCGAAGGAA 1058
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1053 nAspGluAspGluGlyGluProTrpArgArgArgProCysAlaGlyA 1070
   1059 ATCCCGCGTCAGCGACAATTTGCCGATG.....CGGCATACGCCAAAT 1102
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
1070 laProProArgCysCysGlySerProAlaValProArgSerSerAspGly 1086
   1103 ACCCGTCCCTTACCATTCCGGAATAATCGTTCAAACTTGGAGCAGCGT 1152
   :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1087 AlaArgProArgGluAlaPro..... 1093
   1153 TACGGCAAGAAACATCACTCTCAACCGTCGCGCGTCAAAACG 1198
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1094 .....SerSerSerProProThrThrCysArgSerSerThr 1105

```

seq\_name: p1r2:T02345

seq\_documentation\_block:  
hypothetical protein KIAA0324 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 05-Nov-1999  
C:Accession: T02345  
R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;  
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.  
submitted to the EMBL Data Library, March 1998  
A:Description: Sequencing of human chromosome 16p13.3.  
A:Reference number: Z14664

A:Accession: T02345

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1791 <RIC>

A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AA08453.1; PID:g2996650

C:Genetics:

A:Map position: 16

A:Introns: 1610/2; 1706/2

A:Note: KIAA0324

alignment\_scores:

Quality:	150.00	Length:	598
Ratio:	0.558	Gaps:	28
Percent Similarity:	44.983	Percent Identity:	22.742

alignment\_block:

US-09-303-518d-463 x T02345

Align seg 1/1 to: T02345 from: 1 to: 1791

```
39  ACTGGCGAGTGCCT...GCCATGCATGCACACGCTCAGATTGGCAA 85
   ::::::::::: ||| ::::::::::: :::::::::::
611 SerGlySerProGluValAspSerLysSerArgLeuSer...ProAr 626
   ::::::::::: ::::::::::: :::::::::::
86  AGCATCCCTTTATCCGGCAGGTTCTCGACGCTCAGCATTTTCGAACCCGAC 135
   ::::::::::: ::::::::::: :::::::::::
626 gArgSer.....ArgSerGlySerSerProGluValLysAspLysProA 641
   ::::::::::: ::::::::::: :::::::::::
136 GGAATAACCACTATTTCGGCAGCAGGGGGAGCTTGCNAGCGCAACGG 185
   ||| ||| ||| ::::::::::: :::::::::::
641 rGAlaAlaPro.....ArgAlaGlnSerGlySerAspSerProGlu 655
   ::::::::::: ::::::::::: :::::::::::
186 CCATATCGGATTGGGAACATACAAGCCATCAGTTGGCCACCTGATGA 235
   ||| ::::::::::: ||| :::::::::::
656 ProLysAlaProAlaProArgAlaLeuProArgArgSerGlySer 672
   ::::::::::: :::::::::::
236 TTCACAGCGCGCGCTTGAAGGAATATCGGCTACATTGTCGCGCTTTTCC 285
   ::::::::::: ||| |||
672 rSerSerLysGlyArg.....GlyProSerProG 682
   ::::::::::: :::::::::::
286 GATCAGCGGCACAAATTCATTCGCTTCGACAAACCATTCCTCATTC 335
   ::::::::::: :::::::::::
682 LuGlySerSerThrGluSerSerProGluHisProPro..... 695
   ::::::::::: :::::::::::
336 CGATTCTGACGAGCGGTAGTCCGTTACCGGATTCACGCTTTACCGCA 385
   ::::::::::: :::::::::::
696 .....LysSerArgThrAlaArgArgGlySerArgSerProG 709
   ::::::::::: :::::::::::
386 TCATATGGGACGATACGAACACCATCCCGCGGAGGCTATGACGGGCA 435
   ::::::::::: :::::::::::
709 uProLysThrLysSerArgThrProArgArgArgSerArgSerS 726
   ::::::::::: :::::::::::
436 CAGGCGGG.....CGGCTATCCGCTCCCAAGG..... 464
   ::::::::::: ||| ||| |||
726 exProGluLeuThrArgLysAlaArgLeuSerArgSerArgSerAla 742
   ::::::::::: :::::::::::
465 .....CGCAGGAGATATATACAGCTACGACATAAAGCGG 499
   ::::::::::: :::::::::::
743 SerSerSerProGluThrArgSerArgThrProProArgHisArgS 759
```

```
500 TTGCCCAAAATATCCGCTCAACCTGACCGACAA..... 533
   ||| :::::::::::
759 rProSerValSerSerProGluProAlaGluLysSerArgSerSerArgA 776
   :::::::::::
534 .....CCGACGACACCGGACACACGCTTCCCGA... 560
   ||| :::::::::::
776 rArgArgSerAlaSerSerProArgThrLysThrSerArgArgGly 792
   :::::::::::
561 .....CCGTTTCCACAATCCGCGCGCTATCTGCTACGCAAGGAGTAGG... 602
   ||| ||| :::::::::::
793 ArgSerProSerProLysProArgGlyLeuGlnArgSerArgSerArgSe 809
   :::::::::::
603 .CGACGATTCAACACGCGCCACCCG...ATACAGCCCGGAGCTGCACAGA 648
   ||| ||| :::::::::::
809 rArgArgLysThrArgThrThrArgArgArgAspArgSerGly.... 824
   :::::::::::
649 TCGGGCAATGCCCGCGGAGCTTCAACGCGCAGCTGCAGATATCTGCAAAA 698
   ||| ||| ||| :::::::::::
825 .....SerSerGlnSerThrSerArgArgArgGlnArg 835
   :::::::::::
699 CATCATCGCGCGCGCAGGAGAAATTGTCGGCGAGCGGATCCCTGCAGG 748
   :::::::::::
836 SerArgSerArgSerArgValThr...ArgArgArgArgGlyGlySerGl 851
   ||| ||| :::::::::::
749 GTATAAGCGAAGGCTCAACATTTCTGTCATGCACGCTTGGGTCTGCTT 798
   ||| ||| :::::::::::
851 yTyr.....HisSerArg..... 855
   :::::::::::
799 TCCACCGAAACAAAGATGGCGGCATCAACGATTTGGCAGATATGGCGCA 848
   :::::::::::
856 .....SerProAlaArgGlnGluSerSerArgThrSerSer 867
   :::::::::::
849 ACTCAAGACTATGCCGCGCAGCAGCCATCCGCGATTGGGCGAGTCCAAACC 898
   :::::::::::
868 ArgArgArgGlyArgSerArgThrPro.....Pr 878
   :::::::::::
899 CCAATGCGCACAAGCATAGACCGCTCAGCAATATCTTTATGGCAGCC 948
   ||| ||| ||| :::::::::::
878 oThrSerArgLysArgSerArgSerArg.....ThrS 889
   :::::::::::
949 ATCCC.....CATCAAAGGATGGAGCTGTCCGGGGAAATA 986
   ||| :::::::::::
889 erProAlaProTrpLysArgSerArgArgAlaSerProAlaThrHis 905
   :::::::::::
987 CGGCTTGGCGGCATCAGCGCACATCTGTCAACCGCTCCGACAGTGGCG 1036
   ||| ||| :::::::::::
906 ArgArgSerArg.....SerArgThrProLeuLysSerArgArg 919
   :::::::::::
1037 CGATCGCATTCGCCGAAAGGAAATCCCGCTCAGCGACAAATTTGCCGAT 1086
   :::::::::::
919 gSerArgSerArgThrSerProValSerArgArgArgSerArgArgT 936
   :::::::::::
1087 GCGCATACGCCAAATACCGCTCCCTTACCATTCCCGAAATATCCGTT 1136
   :::::::::::
936 hrSerVal..... 938
   :::::::::::
1137 AAATTCGACGCGTTACGGCAAGAAACATCACTCTCTCAACCGTC 1186
   ||| :::::::::::
939 .....ThrArgArgSerArgSerArgAlaSerProVa 950
   :::::::::::
1187 CGCGCTCAACGCGCAAAATGTCAAACTGGCAGACCAACCCACCGCAAG 1236
   :::::::::::
950 lSerArgArgSerArgSerArg.....ThrProProValT 963
   :::::::::::
1237 ACAGCGCTACCGTTTGACGGTAAAGGTTTCCGAATTTTGAGAGCACGT 1286
   ||| ||| :::::::::::
963 hrArgArgArgSer...ArgSerArgThr.ProThrThrArgArgS 978
   :::::::::::
1287 GAAATATGATACGAGCTCGATATTCAGAAATTTATCGGGGGCGGTATAC 1336
   :::::::::::
978 rArgSerArgThrProProValThrArgArgArgSerArgSerArgThrPr 995
   ||| ||| :::::::::::
1337 CTAAGGCTAAGCCTGTGTTTGTATGCGAAACCGAGATGGGAGGTTGATAG 1386
```

```

995 ro.....ProValThrArgArgSerArg.....Ser 1004
1387 AGCTTAATAAATGACAACTCGTAGCAGGTGGAGAAATGTTCCAGG 1436
    : : : : : : : : : : : : : : : : : : : : : : : : : :
1005 ArgThrSerProIleThrArgArgSerArgSerArgThrSerProVa 1021
1437 AACGAGAAGAAGAGTCAGAGTAGTCAGTTAAAGCCCATGCGCAACGAG 1486
    : : : : : : : : : : : : : : : : : : : : : : : : : :
1021 lThrArgArgSerArgSerArgThrSerProValThrArgArgArg. 1037
1487 AATGGGAAATAAACACAGGTTAGATTTTAATCATTTTATAGTGCGTAT 1536
    : : : : : : : : : : : : : : : : : : : : : : : : : :
1038 .....Ser 1038
1537 ATCAATAAGAAAGGCACAGTAACAGGAGGCATAGTCTAACCCGNGGTCA 1586
    : : : : : : : : : : : : : : : : : : : : : : : : : :
1039 ArgSerArgThrSerProValThrArgArgSerArgSerArgThrPr 1055
1587 TGTACGGGTGATACACAAACCTCGSCA.....CTGTATA 1621
    : : : : : : : : : : : : : : : : : : : : : : : : : :
1055 oProAlaIleArgArgSerArgSerArgThrProLeuLeuProArgL 1072
1622 ACATCGGCTTTATCAGGCACAGTGGAAATTAAGCCCTGATGGAAGT 1671
    : : : : : : : : : : : : : : : : : : : : : : : : : :
1072 ysArgSerArgSerArgSerProLeuAlaIleArgArgSerArgSer 1088
1672 TGGGAGTGAAACGAAAGGTTGGGAAAGTCATGACCAAG 1713
    : : : : : : : : : : : : : : : : : : : : : : : : : :
1089 ArgThrProArgThrAlaArgGlyLysArgSerLeuThrArg 1102

```

seq\_name: pir2.JW0067

seq\_documentation\_block:

chitinase (PC 3.2.1.14) A - *Emericella nidulans*

N:Alternate names: ch1A

C:Species: *Emericella nidulans*, *Aspergillus nidulans*

C&gt;Date: 13-Jun-1998 #sequence\_revision 10-Jul-1998 #text\_change 11-Jan-2002

C:Accession: JW0067

R:Takaya, N.; Yamazaki, D.; Horiuchi, H.; Ohta, A.; Takagi, M.

Biosci. Biotechnol. Biochem. 62, 60-65, 1998

A:Title: Cloning and characterization of a chitinase-encoding gene (ch1A) from *Aspergillus*

A:Reference number: JW0067; MUID:98162139

A:Accession: JW0067

A:Molecule type: mRNA

A:Residues: 1-660 &lt;TAK&gt;

C:Cross-references: DDBJ:D87895; NID:q2821948; PID:q2828335

C:Comment: This enzyme hydrolyzes chitin at beta-1,4 bonds between N-acetylglucosamine

C:Genetics:

A:Gene: ch1A

C:Keywords: glycosidase; hydrolase

alignment\_scores:

Quality: 149.00 Length: 399

Ratio: 0.819 Gaps: 19

Percent Similarity: 45.614 Percent Identity: 23.559

alignment\_block:

US-09-303-518D-463 x JW0067 ..

Align seg 1/1 to: JW0067 from: 1 to: 660

97 ATCCGCGCAGGTT.....CTGACCGCTGACGATTTGGAACCGCAGCG 137

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

332 lIleArgLluleuThrAspleuaspProAsnHis.ProProProThrT 348

138 GAAATACCACTATTTCGACGAGGGGGAGCTTGCNAGCCCAACGCGC 187

: :

348 hrSerProThrProThrProThr.....ProSerThrThrThr 360

188 ATATCGGATTGGGAACATACAAAGCCATCATGTTGGGCCACCTGATG 237

||| :

```

361 ThrSerThrThrSerThr..... 366
238 CAACAGCGCGCGTTGAAGAAATATCGCTACATGTCGCGCTTTTCCGA 287
    : : : : : : : : : : : : : : : : : : : : : : : : : :
367 .....ThrSerThrThrSerAlaThrSerT 375
288 TCAGGGCACAATTCATTGCGCCTTCGACAAACCACTGCTCACATTCG 337
    : : : : : : : : : : : : : : : : : : : : : : : : : :
375 hrThrSerThrThrSerThrThrSer..... 386
338 ATTCTGACGAAGCGGTAGTCCCGTTGACGGATTTCACGCTTTACCGCATC 387
    : : : : : : : : : : : : : : : : : : : : : : : : : :
387 .....ThrProThrThrSerThrThrSerThrSe 398
388 CATTGGGACGATACG...AACACCATCCCGCGCAGCGCTATGACGGCC 434
    : : : : : : : : : : : : : : : : : : : : : : : : : :
398 rThrThrThrProThrProSerProSerProSerThrAlaSerSerSerT 415
435 ACAGGGCGGGCTATCCCGTCCCAAGGGCGGAGGATATATACAGCT 484
    : : : : : : : : : : : : : : : : : : : : : : : : : :
415 hrThrGluThrValThrProSerProLysProSer..... 426
485 ACACATAAAGGGGTGCCCAAAATATCCGCTCAACCTGACCGACAAC 534
    : : : : : : : : : : : : : : : : : : : : : : : : : :
427 .....ProSerGluSerSerThrThrSerGluThrSe 437
535 CGCAGCACGGACAAACGGCTTGGCGACCGTTTCCACAATCGCGCGCTAT 584
    : : : : : : : : : : : : : : : : : : : : : : : : : :
437 rSerLeuProSerThrSerThrProValValSerGluThrProSer.... 452
585 GCTGACCAAGGAGTAGCGGAGGATTCAACGCGCCACCGATACAGCC 634
    : : : : : : : : : : : : : : : : : : : : : : : : : :
453 .....GluThrLysThrProThrSerSerAlaProProLeuSerSer 467
635 CCGAGCTGGACAGATCGGGCAATCGCGCGCAAGCTTCAACGGCACTGCA 684
    : : : : : : : : : : : : : : : : : : : : : : : : : :
468 SerSer.....ProValGlySerSerSerTh 477
685 GATATGCTAAAAACATCATCGCGCGGAGAGAAATTCGCGCGCAGG 734
    : : : : : : : : : : : : : : : : : : : : : : : : : :
477 rAlaSerSerThrSerThrProSerGluThrProSerThrSerSerT 494
735 CGATGCGCTGCGGGTATAGCGAAGCTCAACATGCTGCTCATGCGACG 784
    : : : : : : : : : : : : : : : : : : : : : : : : : :
494 hrArg.....AlaValSerGluThrSerThrHisIleSer 505
785 GCTTGGTCTGCTTTCCACCGAAACAGATGCGCGCATCAACGATTG 834
    : : : : : : : : : : : : : : : : : : : : : : : : : :
506 ThrSerThrSerSerGlyProGluThrSerLeuThrGlySerSerThr.. 521
835 GCAGATATGCGCAACTCAAGACTATGCGCGCAGCAGCCATCGCGGATTG 884
    : : : : : : : : : : : : : : : : : : : : : : : : : :
522 ..SerValProAlaThrSerSerValProSer.....SerAlaIleS 536
885 GGCAGTCCAAACCCCAATCGCGCACAAAGCATAGAACGG..... 925
    : : : : : : : : : : : : : : : : : : : : : : : : : :
536 erProSerSerThrProValIleSerGluThrProArgProProValThr 552
926 ...TCAGCAATATCTTTATGCGCCCATCCCATCAAGGGATTGGAGCT 972
    : : : : : : : : : : : : : : : : : : : : : : : : : :
553 SerSerSerSerThrPheValSerSerThrSerThrSerThrSerThr 569
973 GTCCGGGAAAATACGGCTTTGGCGGCATCACGCGCACATCCTGTCAAGCG 1022
    : : : : : : : : : : : : : : : : : : : : : : : : : :
569 sSerGluSerSerThr.....AlaIleGlyThrHis.....SerSerS 582
1023 GTCCACAGATGGCGCGATCGCATTCGCGAAGGAATCCCGCGTCAGCG 1072
    : : : : : : : : : : : : : : : : : : : : : : : : : :
582 erSerIleTip.....GluThrProSerAla 590
1073 ACAATTTTGGCGATCGGCATACGCAAAATACCGTCCCTTACCATTC 1122
    : : : : : : : : : : : : : : : : : : : : : : : : : :
591 Ser.....ThrProAlaSerProSerThrSerPr 601

```







```
185 hrThr..... 186
1037 CGATCGCATTCGCGAAGGAAATCCCGCTCAGCGACAAATTTGCCGAT 1086
187 .....GlyProThrSerProThrThrArgPro.. 195
1087 GCGGCATAGCGCAATACCGTCCCTTACCATTCCGGAATATCCGTTTC 1136
196 .....ProSerThrSerThrProThrSerProThrValProThrS 209
1137 AAACCTTGAGCAGCGTTACGCGAAGAAACATCACCTCTCAACCGTGC 1186
209 erThrThrGluAlaIleThrGlnThrArgLeuSerSerThrThrPro... 224
1187 CGCGCTCAACGCGAATAATGTCAAACTGGCAGACC 1222
225 ThrMetGluThrThrArgThrSerSerTrpGlyThr 236
seq_name: pir2:A48018
```

```
seq_documentation_block:
mucin 7 precursor, salivary - human
N;Alternate names: mucin, MG2; mucin, MG2a-T1; mucin, MG2a-T2; mucin, MG2b-T2
C;Species: Homo sapiens (man)
C;Date: 16-Feb-1994 #sequence revision 18-Nov-1994 #text_change 07-May-1999
C;Accession: A48018; S29115; S29116; S29114
R;Bobek, L.A.; Tsai, H.; Biesbrock, A.R.; Levine, M.J.
J. Biol. Chem. 268, 20563-20569, 1993
A;Title: Molecular cloning, sequence, and specificity of expression of the gene encoding
A;Reference number: A48018; MUID:93388636
A;Accession: A48018
A;Molecule type: mRNA
A;Residues: 1-377 <BOB>
A;Cross-references: GB:113283
A;Experimental source: submandibular gland
A;Note: sequence extracted from NCBI backbone (NCBI:137719, NCBI:137720)
R;Reddy, M.S.; Bobek, L.A.; Haraszthy, G.G.; Biesbrock, A.R.; Levine, M.J.
Biochem. J. 287, 639-643, 1992
A;Title: Structural features of the low-molecular-mass human salivary mucin.
A;Reference number: S29114; MUID:93075006
A;Accession: S29115
A;Molecule type: mRNA
A;Residues: 143-168 <RED>
A;Accession: S29116
A;Molecule type: protein
A;Residues: 5',71-79, 'N',81-86,'XX',89,'X',91,'P' <RE2>
A;Accession: S29114
A;Molecule type: protein
A;Residues: 143-145,'X',147,'XXX',151-152,'X',154-158,'X',160-161,'A',163-164,'XX',167-1
C;Genetics:
A;Gene: GDB:MUC7
A;Cross-references: GDB:138799; OMIM:158375
A;Map position: 4q13-4q21
C;Keywords: glycoprotein
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-377/Product: mucin 7, salivary #status predicted <MAT>
F;97,128,135,146,312/Binding site: carbohydrate (Asn) (covalent) #status predicted
```

```
alignment_scores:
Quality: 147.00 Length: 383
Ratio: 0.987 Gaps: 23
Percent Similarity: 38.903 Percent Identity: 25.326
```

```
alignment_block:
```

```
US-09-303-518d-463 x A48018 ..
```

```
Align seg 1/1 to: A48018 from: 1 to: 377
```

```
119 ACATTTCGAACCCGCGGAAATACCACTATTCGCGCAGCGGGGAG 168
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 SerValAsnProThrLeuValAlaThrThrGlnIleProSerValTh 115
```

```
169 CTTGCCNAGCGCA...ACGGCCATATCGGATTGGGAAACATACAAAGCCA 215
115 rPheProSerAlaSerThrLysIleThr.....ThrLeuProAsnV 129
216 TCAGTTGG.....GCCACCTGATGATTCAACAGCGCGCGTTG 253
129 aThrPheLeuProGlnAsnAlaThrThrIleSerSerArgGluAsnVal 145
254 AGGAAATATCGGCTACATTTGCCGTTTTCGATCAGCGGCGACAAATC 303
146 AsnThrSerSerSerValAlaThrLeuAlaProVal.....AsnSe 159
304 CATTCGCCCT.....TCGACAAACCATGCTCAGATTCGATTCGACGA 347
159 rProAlaProGlnAspThrThrAlaAlaProProThrProSerAlaThrT 176
348 AGCGGTAGTCCCGTTGACGGATTACGCCCTTACCGCATTCATTCGGAGC 397
176 hrProAlaProPro.....SerSerSerAlaProPro 186
398 GATACGAACACCATCCCGCGAGCGGTATGACGGGCCACAGGGCG..... 442
187 GluThrThrAlaAlaProProThrProSerAlaThrThrGlnAlaProPr 203
443 .....CGGGCTATCCGCTCCCAAGG 464
203 oSerSerSerAlaProProGluThrThrAlaAlaProProThrPro.... 218
465 CGCGAGGGATATATACAGTACAGTACATAAAGCGTTGCCCAAAATATCC 514
219 .....ProAlaThrThr.....ProAlaProProSerSer 228
515 GCCTCAACCTGACCGACACCGCAGCGGTATGACGGCTTCGCGACCGT 564
229 SerAlaProProGluThrThrAlaAlaPro.....ProThrPr 241
565 TTCC...ACAATGCCGCGCTATGCTGACGCAAGGAGTAGGCGACGGATT 611
241 oSerAlaThrThrProAla.....ProLeuSerS 251
612 CAACCGCGCGACCGGATACAGCCCGAGCTGGACATCGGCGCAATCGCG 661
251 erSerAlaProProGluThrThr.....AlaValPro 261
662 CGGAAGCTTCA...ACGGCAGTGCAGATATCGTCAAAACATCATCGCG 708
262 ProThrProSerAlaThrThrLeuAspProSerSerAlaSerAla..... 276
709 GCGGCAGGAGAAATTCGCGCGACCGGATGCGTGCAGGGTATAAGCGA 758
276 ..... 276
759 AGGCTCAACATTCGCTGTCATGCAACGGCTTGGCTTCGCTTCCACCGAA 808
277 .....ProProGlu 279
809 ACAAGATGGCGCGCATCAACGATTTCGAGATATGCGCAACTCAAAGAC 858
280 Thr.....ThrAl 282
859 TATGCGCGCAGCAGCCATCCCGATTGGGAGTCCAAACCCCAATCGCGC 908
282 aAlaProProThrProSerAla.....ThrThrProAlaPro. 294
909 ACAAGGCATAGAAGCGTCAAGCAATATCTTTATGGCAGCCATCCCATCA 958
295 .....SerProAla 299
959 AAGGATTCGAGCTCCCGGGGAAATACAGCTTGGCGGSCATCACGGCA 1008
299 ..... 299
```

1009 CATCTGCTCAAGCGGTGGCAGATGGCGGATCGCATTCGCCGAAAGGAA 1058  
 299 .....  
 1059 ATCGCGCGTCAGCGACAATTTTCGGATCGGGCATACGCAATACCCGT 1108  
 300 ..ProGlnGluThrAlaAlaProIleThr...ThrProAsnSerSer 314  
 1109 CCCTTACCAT...CCGAAATATCCGT..... 1135  
 315 ProThrThrLeuAlaProAspThrSerGluThrSerAlaAlaProThrHi 331  
 1136 .CAAACTGGAGCAGGTTCAGCAAGAAACATCACCTCTCTCAACCG 1183  
 331 sGlnThrThrSerValThrThrGlnThrThrLysGlnPro 347  
 seq\_name: pirl:S48478  
 seq\_documentation\_block:  
 N:glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)  
 C:Species: extracellular glucoamylase; mucin-like protein MUC1; protein YIR019c  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 12-Nov-1999  
 C:Accession: S48478; A26877; B26877; S27281; J06123  
 R:Rowley, K.  
 submitted to the EMBL Data Library, October 1994  
 A:Reference number: S48478  
 A:Accession: S48478  
 A:Molecule type: DNA  
 A:Residues: 1-1367 <ROW>  
 A:Cross-references: GB:247047; EMBL:238061; NID:g603997; PID:g763364; GSPDB:GN000009; MIP  
 R:Yamashita, I.; Nakamura, M.; Fukui, S.  
 J. Bacteriol. 169, 2142-2149, 1987  
 A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.  
 A:Reference number: A91831; MUID:87194600  
 A:Accession: A26877  
 A:Molecule type: DNA  
 A:Residues: 1-242 <YAM>  
 A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525  
 A:Accession: B26877  
 A:Molecule type: DNA  
 A:Residues: 762-1331 <YA2>  
 A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526  
 R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
 FEBS Lett. 239, 179-184, 1988  
 A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar  
 A:Reference number: S27281; MUID:89031230  
 A:Accession: S27281  
 A:Molecule type: DNA  
 A:Residues: 1-31 <PAR>  
 A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552  
 R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.  
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
 A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy  
 A:Reference number: JC6123; MUID:96323237  
 A:Accession: JC6123  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1367 <LAM>  
 A:Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387  
 C:Genetics:  
 A:Gene: SGP:MUC1; STA2; MAL5; DEX2; SGD:S0001458  
 A:Cross-references: MIPS:YIR019c; SGD:S0001458  
 A:Map position: 9R  
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
 F:5-21/Domain: transmembrane #status predicted <TM1>  
 F:1350-1366/Domain: transmembrane #status predicted <TM2>

alignment\_scores:  
 Quality: 146.00 Length: 438  
 Ratio: 0.737 Gaps: 16

Percent Similarity: 45.205 Percent Identity: 23.288

alignment\_block:

US-09-303-518D-463 x S48478 ..

Align seg 1/1 to: S48478 from: 1 to: 1367

20 TATCCCTTATTCTGTCCATATCGCAGTGTGCCTGC.....CGATGCAT 63  
 :::: ||| |||||::: ||| :::: |||  
 192 PheProGlyPhe.....TyrTrpAsnIleAspCysAspAsnAsnCysG1 206  
 64 GCACACGCCTCAGATTGGCAACAGATCCCTTTATCCGGCAGGTCTCGA 113  
 :::: ||| :::: |||  
 206 yGlyThr.....LysSerSer 212  
 114 CCGTCAGCATTTTGAACCCGACGCGGAATACCACTATTTCGGCAGCAGG 163  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
 212 hrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 228  
 164 GGGAGCTTGGCCAGCGCAACGCCATATCGGATTGGGAAACATACAAAGC 213  
 :::: ||| :::: |||  
 229 GluSerSerThrThrThrThrThrThrThrThrThrThrThrThrThr 238  
 214 CATCAGTTGGCCACCTGATTTCAACAGCGCGCGCTTCAAGGAAATAT 263  
 :::: ||| :::: |||  
 239 .....GluSerS 241  
 264 CGGTACATTTCCGCTTTCCGATCAGCGGCACAAATTCATTCGCGCCT 313  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
 241 erThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 254  
 314 TCGAACACCATGCTCATCATTCGATTCCTGACG.....AA 348  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
 255 SerThrThrAlaProAlaThrProThrThrThrThrThrThrThrThr 271  
 349 GCCGTAGTCCGTTGACGATTACGCTTTACCGCATCCATTGGGACGG 398  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
 271 sProThrProThrThrThrThrThrThrThrThrThrThrThrThr 282  
 399 ATACGAACACCATCCCGCGG..... 418  
 282 ysGluLysProThrProProHisHisAspThrThrProCysThrLysLys 298  
 419 .....ACGGTATGACGGCCACAGCGCGCGGCGGTATCCCGCTCC 459  
 299 LysThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 315  
 460 AAGGCGCGAGGATATATACAGCTACGACATAAAGGCGTTGCCCAAAA 509  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
 315 oThrProSerSerSerThrThrGluSerSerSerAlaProValProThrP 332  
 510 TATCCGCTCAACCTGACCGCAACACGCGGACGCGGACACGCGTTCGCG 559  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
 332 roSerSerSerThrThrGluSerSerSerAlaProValThrThrThrThr 348  
 560 ACCGTTTCCACAATGCGCGCTATGCTGACGCAAGGAGTAGCGCAGGA 609  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
 349 ThrGluSerSerSerAlaProValProThrProSerSerSerThrThrG1 365  
 610 TTCAACCGCCCGCGGATACACCGCGAGCTGACAGATCGGCGCAATGC 659  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
 365 uSerSerSerAlaProValThrSer...SerThrThrGluSerSerSerA 381  
 660 CGCGAAGCGCTTCAAGCGGCACTGCAGATATCGTCA.....AAA 697  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
 381 laProValThrThrThrThrThrThrThrThrThrThrThrThrThrThr 397  
 698 ACATCATCGCGCGGAGGAGAAATTGTGCGCGCAGCGGCGGCGTGCAG 747  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
 398 ProSerSerSerThrThrGluSerSerSerAlaProValThrThrThr 414  
 748 GGTATAGCGAAGGCTCAACATTTGCTGTCATCGACGCGTTGGGTCTGCT 797  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: |||

414 rThrGluSerSerAlaProValThrSerThrThrGluSerSers 431  
 798 TTCCACCGAAACACAGATGCGCGCATCAAGATTGGCAGATATGGCGC 847  
 431 erAlaProValThr.....SerSerThr.....Thr 439  
 848 ACTCAAGACATATCCCGCAGCAGCATCCGCGATTGGCAGTCCAAAC 897  
 440 GluSerSerAlaProValThrSerThrThrGluSerSerAla 456  
 898 CCCATGCGGCACAGGCATAGAACCGCTCAGCAATA.....TCCT 938  
 456 aProValPro.....ThrProSerSerThrThrGluSers 469  
 939 TATGCGAGCATCCCATCAAGGATGGAGCTGTCGCGGGAATAACG 988  
 469 erSerAlaProValThrSerThrThrGluSerSerAlaProVal 485  
 989 GCTTGGCGGATCAGGCACATCTGTCAAGCGTGCAGATGGCGCG 1038  
 486 ProThrProSerSerThrThrGluSerSerAlaProValThrSe 502  
 1039 ATCGCATTCGCGAAAGGAATCCGCG.....TCAG 1070  
 502 rSerThrThrGluSerSerAlaProValProThrProSerSerSert 519  
 1071 CGACAATTTGCGGATGCGGCATACGCCAAATACCGTCCCTTACCATT 1120  
 519 hrThrGluSerSerAlaProAlaProThrProSerSerThrThr 535  
 1121 CCCGAATATCGTTCAACTGGAGCGGTAGCGCAAGAAACATC 1170  
 536 GluSerSerAlaProValThrSerThrThrGluSerSerAla 552  
 1171 ACCT...CCTCAACGTCGCGCGCAACGCAAAATGTCAACTGGC 1217  
 552 aProValProThrProSerSerThrThrGluSerSerThrProv 569  
 1218 AGACCAACGCCACC 1231  
 569 alThrSerThr 573  
 seq\_name: pir2:I47141  
 seq\_documentation\_block:  
 gastric mucin (clone PGM-2A) - pig (fragment)  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 03-Nov-2000  
 C:Accession: I47141; S55315  
 R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.  
 Gastroenterology 106, 200, 1994  
 A:Title: Pig gastric mucin: isolation and characterization of a cDNA clone with a novel  
 A:Reference number: I47141; MUID:94102478  
 A:Accession: I47141  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-528 <TUT>  
 A:Cross-references: EMBL:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208  
 R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.  
 Biochem. J. 308, 89-96, 1995  
 A:Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.  
 A:Reference number: S55315; MUID:95275264  
 A:Accession: S55315  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-528 <TUT>  
 A:Cross-references: GB:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208  
 C:Superfamily: pig submaxillary mucin

alignment\_scores:  
 Quality: 145.50 Length: 479  
 Ratio: 0.591 Gaps: 21

Percent Similarity: 51.357 Percent Identity: 23.173  
 alignment\_block:  
 US-09-303-518d-463 x I47141 ..  
 Align seg 1/1 to: I47141 from: 1 to: 528  
 92 CCTTTATCCGCGAGTTCGACCGTCAGCATTCGAACCCGAGGAAA 141  
 1 ProlSerValGlnProSerSer...SerSerSerProThrThrSe 16  
 142 TACCAC.....TATTCGCGAGCGGGGAGCTTGCNAGCG 179  
 16 rThrThrSerValGlnSerSerSerSerValProIleProSert 33  
 180 CAACGGCCATA.....TCGGATTGGGAAACATACAAAGCC 214  
 33 hrThrSerValGlnProSerSerSerGlySerAlaProThrThrSe 49  
 215 ATCAGTTGGGCCACCTCATGATTCAACAGCGCGCTTGAAGGAATATC 264  
 50 ThrSerValGlnThrSerSerSerSerProPro.....IleSe 63  
 265 GGCTACATTG.....TCCGCTTTTCGATCACGG 293  
 63 rSerThrIleSerValGlnThrSerSerSerSerValProThrThrS 80  
 294 GCACAAATTCATTCGCTTCGACACACCATGCTCACATTCGGATTCTG 343  
 80 erThrThrSerValGlnProSerSerSer.....AlaProThr 94  
 344 ACGAAGCGGTAGTCCGTTGACGGATTACGCTTTACGCGCATCTGG 393  
 95 ThrArgAlaThrSerValGlnSerSerSerSerSerAlaProIleSe 111  
 394 GACGGATACGAACACCATCCCGCGCGGCTATGACGGCCACAGGCGG 443  
 111 rSerThrThrSerValGlnProSerSer.....SerG 122  
 444 CGGCTATCCCGCTCCCAAGCGCGGAGGATATATACGCTACGACATAA 493  
 122 lySerValProThrThrSerAlaThrSerValGlnSerSerSer 138  
 494 AAGCGTTGCCCCAAAATATCCGCTCAACGACGACACCGCAGCACCC 543  
 139 SerAlaProThrThrSerAlaThrSerValGlnProSerSerSerSe 155  
 544 GGCAACGGCTTCCGACCGCTTTCCAAATGCGCGGCTATGCTGACGCA 593  
 155 rProProIleSerSerThrThrValSerValGlnPro..... 166  
 594 AGGATAGGCGCGGATTCAACGCGCGCCCGATACAGCCCGAGCTGG 643  
 167 .....SerSerSerSerAlaProThrThrSerAlaThrSerVal 180  
 644 .....ACAGATCGGCAATGCGCGGAGGCTTCAACG...GCACTGCA 684  
 181 GlnProSerSerSerSerProProIleSerSerThrValSerValG 197  
 685 GATATCGTCAAAA.....ACATCATCGCGCGGCGAGAGAGA 719  
 197 nThrSerSerSerSerValProThrThrSerThrThrSerValGlnP 214  
 720 AATTGTCGCGCGAGGCGATGCGG.....TCGAGGGTA 751  
 214 roSerSerSerSerValProThrThrSerAlaThrSerValArgSer 230  
 752 TAAGCAAGGCTCAACATTGCTCATCGACGGCTTGGTCTGCTTTCC 801  
 231 SerSerSerSerThrProIleProSerThrThrSerVal.....G 245  
 802 ACCGAAACAGATGCGCGGCATCAACAGATTGGCAGATATGCGCACT 851  
 111:..... 111 111



```

738 .....TCCGTCGAGGTTATAGCGAAGGCTCAAA 767
    ||| |||||
440 roThrLeuAspProGlyProCysAlaAlaGly..... 450
768 CATTCCTCTCATGACGGCTGGGTCTGCTTCCACCGAAACAAAGATGG 817
450 ..... 450
818 CGCGCATCAACGATTGGCAGATATGGCGCAACTCAAGACTATGCCGCA 867
    ||| ||||| |||||
451 .....ArgSerGlyArgProGlyAla.....AlaArgS 460
868 GCAGCCATCCCGATTGGCAGTCCAAACCCCAATCCGCGCACAAAGCAT 917
    ||||| ||| |||||
460 erAspHis.....ArgAspArgHis 466
918 AGAAGCCGT.....CAGCAATATCTTTA 940
    ||| |||
467 ArgHisArgAlaGlyHisArgArgHisThrProAlaAlaHisAlaAr 483
941 TGGCAGGCAT.....CCCATCAAGGATTGGAGCTGTCGGGGA 981
    ||||| ||||| |||||
483 gGlyThrHisasnGlyGlyProHisGlnGlnSer.....GlyP 496
982 AAATACGGCTTGGCGGCA.....TCACGGCACATCTCTCAAGCGGT 1024
    ||| ||||| |||||
496 roAlaArGArgGlyArgProGlyAlaAlaArgHisArgAlaSerGly 512
1025 CGCAGATGGCGGATCGCATTCGCGAAAGGAAATCCCGCGTCAGGGAC 1074
    |||||
513 GlnArg..... 514
1075 AATTTTGGCGATGGCGATACGCAATACC.....GCTCCCTTACCA 1118
    ||||| ||||| |||||
515 .....ArgHisThrProArgSerGlyArgArgProLeuArgL 527
1119 TTCCC 1123
    |||
527 euPro 528
seq_name: pir2:B47328

```

```

seq_documentation_block:
natural killer cell tumor-recognition protein - mouse
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumor
C:Species: Mus musculus (house mouse)
C>Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 05-Nov-1999
C:Accession: B47328; I77662
R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A:Title: A cyclophilin-related protein involved in the function of natural killer cells.
A:Reference number: A47328; MUID:93133824
A:Accession: B47328
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1507 <AND>
A:Cross-references: GB:L04289; NID:g192866
A:Note: authors translated the codon AGT for residue 972 as Arg
R:Rinfret, A.; Anderson, S.K.
Mol. Immunol. 30, 1307-1313, 1993
A:Title: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicing m
A:Reference number: I57820; MUID:94019422
A:Accession: I77662
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 235-237; 263-294 <RIN>
A:Cross-references: GB:S65998; NID:g425701; PIDN:AAB28500.1; PID:g425702
C:Genetics:
A:Gene: NK-TR
C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C:Keywords: alternative splicing; lymphocyte
F:60-230/Domain: cyclophilin homology <CYP>

```

```

alignment_scores:
  Quality: 143.50      Length: 507
  Ratio: 0.595        Gaps: 29
  Percent Similarity: 47.535  Percent Identity: 25.049

alignment_block:
US-09-303-518D-463 x B47328 ..
Align seg 1/1 to: B47328 from: 1 to: 1507

304 CATTCGCCCTTCGACAACCAT.....GCCTCACATTCGATTCTGACGA 347
    ||||| |||:||||| ||||| |||||
567 HisSerSerArgAspSerTyrArgSerLysSerHisSerArgSerAspSe 583
348 AGCGCGTAG..TCCGTTGACGGATTACGCTTTACGCGCATCATTGGGAC 396
    : ||| ||||| |||
583 rArgGlySerSerArgSerArg.....A 591
397 GGATAGCAACACATCCCGCCGACGGCTATGACGGCCACAGGGCGGCG 446
    :||: ||| ||| ||| |||: |||
591 laValSerLysSerSerSerArgSerLeuAsnArgSerLysSerArgSer 607
447 CTATCCCGCTCCCAAGGCGGAGGATATATACAGCTACGACATAAAG 496
    ||||| |||: |||: |||:
608 SerSerArgSerGlyProArgArgThrSerIleSer..... 619
497 GCGTTGCCCAAAATATCCGCTCAACCTGACGACACACGCGACCGGA 546
    ||||| ||| ||| |||: |||
620 .....ProLysLysProAlaGlnLeuSerGluAsnLysProValLys 634
547 CAACGGCTTGGCGACCG...TTTCACATATGCGGGCGCTATGCTCAGCG 593
    || ||||| ||||| ||| |||||
634 hrGluProLeuArgProSerValProGln.AsnGlyAsnValLeuValG 650
594 AGGAGTAGCGGACGGAT.....TCAAACGCGCCA 622
    | ||| |||||: |||: |||: |||
650 nProVal..AlaAlaGluAsnIleProValIleProLeuSerAspSerPro 666
623 CCCGATACAGCCCGGACG...TGGACAGATCGGGCAATGCCGCCGAGGCC 669
    ||| ||||| |||
667 Pro.....ProSerArgTrpLysProGlyGlnLysProTrpLysPr 680
670 TTCA.....ACGGCACATGCAGATATCTCAAAACA..... 700
    ||||| |||: ||||| |||||
680 oSerTyrGluArgIleGlnGluMetLysAlaLysThrHisLeuLeuP 697
701 .....TCATCGCGCGCGGAGGAGAAATTCTCGCGCAGCGCATGCCG 742
    |||||: |||: ||||| |||||
697 roValGlnSerThrTyrSerLeuThrAsnIleLysAlaThrVal..SerSe 713
743 TCGAGGGTATAGCGAAGGCTCAAAATGCTGTCATGACGCGCTTGGGT 792
    :|||: ||| ||||| |||
713 rSerSerTyrHisLysArgGluLysPro.....SerGluSerAspGlyS 728
793 CTGCTTTCCACGGAACAAAGATCGCGCATCAACGATTTGGCAGATAT 842
    |||||: |||: ||||| |||||
728 erAlaTyrSerLysTyrSerAspArgSerSerGlySerSerGlyArgSer 744
843 GCGCAACTCAAGACTATGCGCGACGACCATCCGCGATTGGGCGAGTCC 892
    |||||: |||: ||||| |||||
745 GlySerLysSerSerArgSerArgSer..... 754
893 AAAACCCCAATGCCGCACAAAGCATAGAGCGCTCAGCAATATCTTTATG 942
    |||||: |||: ||||| |||||
755 .....SerArgSerTyrThrArgSerArgSerArgSerLeuProT 768
943 GCAGGCAT.....CCCATCAAGGAGGATTGGAGCTGTCGCG 977
    |||||: |||: ||||| |||||
768 hrSerArgSerLeuSerArgSerProSerSerArgSerHisSer...Pro 783
978 GGGAAATACGGCTTGGCGGCATCAGCGCACATCTGTCTCAAGCGGTCCG 1027
    :||| ||| |||: ||| |||: |||:

```





```

648 ATCGGGCAATGCGCGCCGAAGCCTTCAACGGCACTCGCAGATATCGTCAAAA 697
   :: :: |||||::: ||::: |||
943 luSerSerProProSerThrSerValIleLeuThrThrMetAlaSer 959
   ACATCATCGCGCGCGCAGGAGAAATGTTCGGCGCAGGCGATCCGCTGCAG 747
960 ThrGluGlyThrSerGlyAspThrGlyHisThrThrAlaValThrAspG1 976
748 GGTATAACGGAAGGCTCAAAACATTCGTGTCATCGCAGCGCTGGTCTGCT 797
976 nGlySerThrProAlaThrThrGluIleSerValThr ..... 988
798 TTCACCCGAAAAGATGCGCGCATCAACGATTTCGCAGATATCGCGC 847
   |||:::||||::: :: |||||::: ::
989 .....ProSerThrGln...LysMetSerThrValSerThrLeuValThr 1002
848 AACTCAAAAG .....ACTATGCGCGCAGCAGCCATCCGCGGATTGG 895
1003 SerThrGlnGluLeuThrSerSerGlnSerGlnArgThrGlySerMetG1 1019
886 CGAGTCCAAAACCCCAATGTCGCGCACAGGCATAGAACCGCTCAGCAATAT 935
   | ||::: ||| :::: |||||::: ::
1019 yThrSerSerLysPro.....GlnAlaThrThrProThrGluValT 1033
936 CTTTATGCGAGCCATCCCATCAAGGGATTCGAGCTGTCCGGGGAAAAAT 985
   ::
1033 hr.....:::..ThrSer 1035
986 ACGGCTTGGCGGCATCACGGCACATCTCTCAAGCGGTGCGAGATGGGC 1035
   ||| ::: ||||| ||| |||
1036 ThrLeuSerSerPheSerArgGlySerLeuPheSer.....Al 1048
1036 GCGATTCGATTGC.....CGAAGGGAATCCGCGTCAGCGACACAA 1076
   ||::: :: ||| :::: |||||::: ::
1048 aArgAsnScyScysLeuGlnThrLysLysProLeuProAlaValValC 1065
1077 TTTTCCCG .....ATCGGCATACGCCAATACC 1105
1065 ysLeuProAspProSerSerValProSerLeuMethHisSerSerLysPro 1081
1106 CGTCCCTTACCATCCCGAAATATCCGTTCAAACTTGGCAGCGCTTAC 1155
   |||::: ||| :::: |||||::: |||||:::
1082 GlnAlaThrThrProThrGluValThrThrSerThrLeuSerSerPheSe 1098
1156 GCGAAAGAAACATCACCTCTCAACCGTGGCGCGCTCAACGGCGAAAAA 1205
   :: ::|||
1098 rArgGlySerThr.....GlnThrGlnThrV 1107
1206 TGTCAAACTGGCGAGACCAAGCCACCGCAAGACAGCGGTACCGTTTGACG 1255
   ::||| |||||:::||||::: ::: ::: |||
1107 aiSer...TrpGluThrSerSerSerGlyLysIleThrAlaProSerThr 1122
1256 GTAAGGGTTTCCGAATTTTGAGAAGCAGCTGAAATATATGATACGAAGCTC 1305
   ||| ||| :::::
1123 SerSer.....ArgArgThrProSerValAlaThrSerAs 1134
1306 GATATTC 1312
   :|||||
1134 pilePhe 1136
seq_name: pir2:T38459

```

seq\_documentation\_block:  
hypothetical divergent repeat-containing protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
C:Accession: T38459; T38380  
R:Harris, D.; McDonald, S.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
Submitted to the EMBL data Library, February 1996  
A:Reference number: Z21794  
A:Accession: T38459  
A:Molecule type: DNA

A; Residues: 1-1748 <HAR>  
A; Cross-references: EMBL:Z69368; PIDN:CAA93290.1; GSPDB:GN00066; SPDB:SPAC27F1.01c  
R;McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.  
submitted to the EMBL Data Library, September 1997  
A;Reference number: Z21790  
A;Accession: T38380  
A;Molecule type: DNA  
A;Residues: 1457-1794 <MCL>  
A; Cross-references: EMBL:Z70691; PIDN:CAA94638.1; GSPDB:GN00066; SPDB:SPAC25G10.09c  
A;Experimental source: clone c25G10  
C;Genetics:  
A;Gene: SPAC27F1.01c; SPDB:SPAC25G10.09c  
A;Map position: 1

```

alignment_scores:
  Quality: 143.50      Length: 420
  Ratio: 0.829        Gaps: 20
  Percent Similarity: 41.190  Percent Identity: 22.143

alignment_block:
  US-09-303-518D-463 x T38459  ..

Align seg 1/1 to: T38459  from: 1  to: 1794

112 CGGAAAATATACCCCTTATTCTTCATCTGGCAGTGTGCGCTGCCGATGC 61
|||||:|||||  |||  ::|||::|  ::|||  ::|||  ::|||
1326 ProSerThrThrSerThrSerPheAsnThrAlaProIleProGlnGlnAl 1342
62 ATGCACAGCCCTCAGATTGGCAAGAGTCCCTTTATCCGGCAGGTCTC 111
|  ::  ::  ::|||::|||  |||||  |||
1342 aProLeuGluAsnGlnPheSerLysMetSerLeuGluPro..... 1355
112 GACCGTCAAGATTTCGAACCCGACGGGAAATACCACTATTTCGGCAGCAG 161
1355 ..... 1355
162 GGGGAGCTTGGCCNAGCGCAACGGCCATATCGGATTTGGGAAACATACAA 211
|||||  |||  |||  ::  |||:
1356 .....ProValArgProAlaValProThrSe 1364
212 GCATCAGTTGGCCACCTGATGATTCACACAGCGGCCGCTTGAGAGAAAT 261
|||||:  |||||::|||
1364 rProLysProGlnIleProAspSerSer..... 1373
262 ATCGGCTACATTGTCCGGCTTTCCGATCACGGGCACAAATTCATTCGCC 311
::|::|::|::|::|  |||  ::|::|::|
1374 ..AsnValHisAlaProPro..Pro.....ProValGlnP 1384
312 CTTTCGACACCACTAGCT...CACATTCGATTCTGACGAAGCCGGTAGTC 358
||  ::|::|::|::|  |||  ::  ::|::|::|  ::
1384 roMerAsnAlaMetProSerHisAsnAlaValAsnAlaArgProSerAla 1400
359 CGGTTTCAGCGATTCAGCCCTTT.....ACCGC 384
|||  ::|::|::|::|
1401 ProGluArgArgAspSerPheGlySerValSerSerGlySerAsnValSe 1417
385 ATCCATTGGGCGGATACGACCAACCACTCCG..... 415
:|||||  ::|::|::|::|::|::|
1417 rSerIleGluAspGluThrSerThrMetProLeuLysAlaSerGlnPro 1434
416 .....CCGACGGCTATGACGGGCCACAGCGCGCGCTATCCCG..... 454
|||  |||  ::|::|::|::|  ::|::|::|
1434 hrAsnProGlyAlaProSerAsnHisAlaProGlnValValProProAla 1450
455 .....CTCCCAAGGCGCGGAGGATATATAC 480
|||||  |||::|  |||::|
1451 ProMetHisAlaValAlaProValGlnProLysAlaProGlyMetValTh 1467
481 AGCTACGACATAAAGCGGCTTGCCCAAAATATCCGCGCTCAACCTGACCGA 530
|  ::  ::
1467 rAsnAlaPro.....AlaProSerSerAlaProA 1477

```

531 CAACGCGACGACGACGCGCTTCCGACCGCTT ..... 565  
1477 laProAlaProValSerGlnLeuProProAlaValProAsnValPro 1493  
566 .....TCCACATGCGCGCGTATGCTGACGCAAGGAGTAGGCGACGGA 609  
1494 ValProSerMetileProSerValAlaGlnGln .....ProPr 1506  
610 TTCAAACGGCGCCACCGATACAGCCCGGAGCTGGACAGATCGGGCAATGC 659  
1506 oSerSerValAlaProAlaThrAlaProSerSerThr .....Leup 1520  
660 CGCGGAAGCCTTCAACGCGCACTGCAGATATCGTCAAAACATCATCGCGG 709  
1520 roProSerGlnSerPheAlaHisValProSerProAla ..... 1533  
710 CGCGAGGAGAAATGTGCGCGGAGGCGATGCGTGCAGGGTATAAGCGAA 759  
1533 ..... 1533  
760 GGCTAAACATGCTGCTCATGACGCGCTTGGGTCTGTTTCCACCGAAA 809  
1534 .....ProPro ..... 1535  
810 CAGATGGCGCGCATCAACGATTTGCCAGATATGCGGCAACTCAAGACT 859  
1535 ..... 1535  
860 ATGCCGAGCAGCATCCGCGATTGGGAGTCCAAAACCCCA ..... 901  
1536 AlaProGlnHisProSerAlaAlaLeuSerSerAlaProAlaAspAs 1552  
902 ....ATGCCGCAACAGGCA .....TAGA 920  
1552 nSerMetProHisArgSerSerProTyrAlaProGlnGluProValGlnL 1569  
921 AGCGCTCAGCAATATCTTATGGCAGCATCCCATCAAGGAGTGGAG 970  
1569 ysProGlnAlaIleAsnAsnIleAlaProAlaThrAsnLeuGlyThrSer 1585  
971 CTGTCGGGGAAAATACGCGCTTGGG .....CGGCATCAGCGCACATCCT 1014  
1586 GlnSerPheSerPro.ArgMetGlyProValAsnAsnSerGlySer.Pro 1601  
1015 GTCACGCGTCCGAGATGGGC .....GGATCGCATTTGCCGAAAGGAA 1058  
1602 LeuAlaMetAsnAlaAlaGlyGlnProSerLeuAlaValProAlaValPr 1618  
1059 ATCCGCGCTCAGGACAAATTTGCGGATCGCGCATACGCCAAATACCGGT 1108  
1618 oSerAlaProSerAsnHisPheAsnProPheAlaLysMetGlnProProA 1635  
1109 CCCT 1113  
1635 laPro 1636

seq\_name: pir2.T29634

seq\_documentation\_block:

hypochemical protein C12D12.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000

C:Accession: T29634

R:Nhan, M.; Hawkins, J.

submitted to the EMBL Data Library, March 1996

A:Description: The sequence of C. elegans cosmid C12D12.

A:Reference number: 220656

A:Accession: T29634

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-825 &lt;NHA&gt;

A:Cross-references: EMBL:U51998; PIDN:AAA96080.1; GSPDB:GN00028; CESP:C12D12.1

A:Experimental source: strain Bristol N2; clone C12D12

C:Genetics:

A:Gene: CESP:C12D12.1

A:Map position: X

A:Introns: 48/1; 86/3; 137/1; 172/3; 224/3; 253/1; 287/3; 328/2; 454/1; 487/3; 692/1

C:Superfamily: Epstein-Barr virus membrane antigen gp350

alignment\_scores:

Quality: 143.00 Length: 376

Ratio: 0.851 Gaps: 16

Percent Similarity: 44.681 Percent Identity: 22.074

alignment\_block:

US-09-303-518D-463 x T29634 ..

Align seg 1/1 to: T29634 from: 1 to: 825

110 TCGACCGCTCAGCATTTCAACCCCGACGGGAATACACCTATTTCGGCAGC 159  
473 ThrValSerThrMetSerProThrThrValThrValProThrTh 489  
160 AGGGGGAGCTTGCNAGCGCACGCGCATATCGGATTGGGAAACATACA 209  
489 rProThrProValProThrThr ..... 497  
210 AAGCCATCAGTTGGCCACCCTGATGATTCACAGCGCGCGCTTGAAGAA 259  
498 .....AsnThrProProAlaAsnPro 504  
260 ATATCGCTACATTTCCGCTTTTCCGATCAGCGGCACAAATTCCTATCG 309  
505 ThrThrAlaThr .....ProThrThrValGlyThrSerLysGI 517  
310 CCCTTCGACAAACCATGCTCAGATTCCGATTCTGACGAAGCGGTAGTCC 359  
517 nThrAsnThrIleSerProHisLeuSerThrIleThrGlySerIleVal. 533  
360 CGTTGACGAGTTCAGCCTTTACCGCATTCATGGAGCGATACGACACAC 409  
534 .....ThrSerThrProThrMetAlaProGlnThrSerAla 545  
410 ATCCCGCGGAGCGGTATGACGGCCACAGCGCGCGCTATCCGCTCCC 459  
546 SerProThrThrThrProThr...HisThrAlaSerGlnProThrTh 561  
460 AAGGCGCGAGGATATATACAGTACGACATAAAGGCGCTGCCAAAAA 509  
561 rThrLysPro .....ValValThrThrAsnSerValThrProSerT 575  
510 TATCCGCTCAACCTGACCG .....ACAACCGCAGCACCGGACAAAC 550  
575 hrGlyThrThrThrValProValProThrThrThrGlySerPro ..... 589  
551 GGCTTGGCGACCGTTTCCACAATGCCGCGCTATGCTGACGCAAGGAGTA 600  
590 .....ThrGlnThrThrAlaProValThrLysProThrValProse 604  
601 GCGACGATTCAACCGCGCCAGATACAGCCCGGAGCTGGACAGATC 650  
604 rSerThrThrGlnThrAlaProProValThrThrProThr ..... 617  
651 GGGCAATGCCCGGAGCGCTTCAACGCGCATCGCATATCGTCAAAAACA 700  
618 ..SerGlnProProValThrThrSerLeuLeuThr ..... 630  
701 TCATCGCGCGGAGGAGAAATGTGCGCGCAGCGCATGCCGTGCAGGGT 750  
631 .....LeuThrThrProThrValProVal ..... 638  
751 ATAAGCGAAGGCTCAAAATTTGCTGTGTCATGACGCGCTTGGGTC...TGCT 797  
639 .....ThrThrThrValValProSerSerSerAlaThrValProThrT 652

```
798 TTCACCGCAAAACAGATGGCGCGCATCAACAGATTGGCAGATATGGCGC 847
||||| ||| ||||| |||
652 hrProProThrThrValThrValAlaAlaThrThrSer..... 665
848 AACTCAAGACTATCGCGCAGCAGCCATCGCGGATGGCGAGTCCAAAC 897
||||| ||||| ||||| |||
666 LysAlaProValThrThrSerProThrLeuAlaProThrSerProTh 682
898 CCCATGCCGACACAGCAGATAGACCGTCAGCAATATCTTATGGCAGC 947
||||| ||||| ||||| |||
682 rLysLeuProThrSer.....ProProSerThrValGly.....T 694
948 CATCCCATCAAGGATGGAGTGTCCGGGAAATACGGTTGGCGC 997
||||| ||||| ||||| |||
694 hrSerProThrAlaProAlaAsnLeuThrThrProThrAla..... 708
998 GCATCAGGCACATCCTGTCAAGCGGTGCGCAGATGGCGCGATGCGCAT 1047
708 ..... 708
1048 CCGAAGGGAATCGCGTCAGCGACAATTTGCCGATGGCGCATACGC 1097
||||| ||||| ||||| |||
709 ....ProValAsnProThrSerThrThrAlaProThrAlaProValA 724
1098 CAATATACCGTCCCGTACCATTCGCGAAATA..... 1129
||| ||| ||| ||| |||
724 snProThrSerProThrThrAlaProThrValProProValThrThr 740
1130 .....TCCGTTCAAACTTGGAGCAGCGTTAC 1155
||||| ||||| ||||| |||
741 ThrProThrThrThrSerThrThrThrThrThrThrThrThrThr 757
1156 GCGAAGAAACATCACCTCCTCACCG 1183
||||| ||||| ||||| |||
757 rThrThrGlnThrThrProThrThrPro 766
seq_name: pir2:T47182
```

seq\_documentation\_block:  
hypothetical protein DKFZp434M1616.1 - human (fragment)

C:Species: Homo sapiens (man)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47182  
R:Koehler, K.; Beyer, A.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: Z24378  
A:Accession: T47182  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1299 <AAA>  
A:Cross-references: EMBL:AL162004  
A:Experimental source: adult testis; clone DKFZp434M1616  
C:Genetics:  
A:Note: DKFZp434M1616.1

alignment\_scores:  
Quality: 142.00 Length: 587  
Ratio: 0.568 Gaps: 29  
Percent Similarity: 42.589 Percent Identity: 21.465

alignment\_block:

```
US-09-303-518d-463 x T47182 ..
Align seg 1/1 to: T47182 from: 1 to: 1299
74 CAGATTGG.....CAAACGAT 90
|||||
254 GlnValTrpAsnLysLysAsnAlaAsnGluLysGlyArgSerGlnThrSe 270
91 CCCTTTATCCGCGAGGTCTCGACCGTCAGCATTTCAACCCGCGGAA 140
||| |||||
270 rLysLeuProProArgPheAlaLys.....LysGlnAlaThrGlyI 284
```

```
141 ATACCACCTATTCCGACAGGGGGAGCTTCCNAGCGCAACGGCCATA 190
||||| ||||| ||||| |||
284 leGlnAlaGlnSerSerAlaSerValProProLeuAlaSerAlaPro 300
191 TCGGATTGGGAACATACAAAGCCATCAGTTGGGCCACCTGATTTCAA 240
||||| ||||| ||||| |||
301 LeuProProSerThrSerAlaSerValProAlaSerThr..... 313
241 CAGGCGCGCTTGAAGAAATATCGGCTACATTTCCGCTTTTCCGATCA 290
||||| ||||| ||||| |||
314 .SerAlaProLeu.....ProAlaThrLeuThrProValProAla 327
291 CGGGCACAAATTCCTCCCTCCGACCAACCATGCTCACATT..... 334
||||| ||||| ||||| |||
327 erThrSerAlaProValProAlaSerThrLeuAlaProValLeuAlaSer 343
335 .....CCGATTCTGACGAGCGGTAGTCCCGTTGACGGATTACG 375
||||| ||||| ||||| |||
344 ThrSerAlaProValProAlaSerProLeuAlaProValSerAlaSerAl 360
376 C...TTTACCGCATTCATTGGGACGATACGAACACCATCCCGCCGACG 422
||||| ||||| ||||| |||
360 aSerValSerAlaSerValProAlaSerThrSerAlaAlaIleThrS 377
423 CTATGACGGCCACAGGGCGGCTATCCCGTCCCAAAGCGCGAGGG 472
||||| ||||| ||||| |||
377 erSerSerAlaProAlaSerAlaProAlaProThrProIleLeuAlaSer 393
473 ATATATACAGCTACGACATAAAGGGGTTGCCCAAAATATCC..... 514
||||| ||||| ||||| |||
394 ValSerThrProAlaSerValThrIleLeuAlaSerAlaSerIleProI 410
515 .....GCCTCAACCTGACGGCAACACCGCAGCAGCGGACA 548
||||| ||||| ||||| |||
410 eLeuAlaSerAlaLeuAlaSerThrSerAlaProThrProAlaProAla 427
549 ACGGCTTGGCG.....ACGTTTCCACAATGCCGGCGCTA 583
||||| ||||| ||||| |||
427 laSerSerProAlaAlaProValIleThrAlaProThrIleProAlaSer 443
584 TGTGACGCAAGAGTAGGCGGATTCAAACGCCCA..... 622
||||| ||||| ||||| |||
444 .....AlaProThrAlaSerValProLeuAlaProAl 454
623 .....CCGATACAGCCCGAGCTGGACAGATCGGCAATGC 659
||||| ||||| ||||| |||
454 aSerAlaSerAlaProAlaProAlaProThrProValSerAlaProAsn 471
660 CGCCGAAGCCTTCAAGCGCACTGCAGATATCGTCAAAACATCATCGCG 709
||||| ||||| ||||| |||
471 roAlaProProAlaProAla...GlnThrGlnAlaGlnThrHisLysPro 486
710 CGGCAGAGAAATTG.....TCGGCGCAGCGGATGCC 741
||||| ||||| ||||| |||
487 AlaGlnAsnProLeuGlnThrThrSerGlnSerSerLysGlnProPro 503
742 GTGAGGGTATACGCAAGGCTCAACATTCGTCATCGACCGCTTGGG 791
503 oSerIleArgLeuProSerAlaGlnThr..... 512
792 TCTGCTTTCACCGCAAAACAGATGGCGCGCATCA..... 826
||||| ||||| ||||| |||
513 .....ProAsnGlyThrAspTy-ValAlaSerGlyLysSerIleGln 526
827 .....ACGATTGGCAGATATGGCGCAACTCAA 855
||||| ||||| ||||| |||
527 ThrProGlnSerHisGlyThrLeuThrAlaGluLeuTrpAspAsn...Ly 542
856 GACTATGCCCGCAGCAGCCATCCCGGATGGGCAG...TCCAAACCCCAA 902
||||| ||||| ||||| |||
542 sValAlaPro.....ProAlaValLeuAsnAspIleSerLysLysLeuG 557
```

```
903 TGGCGCAGCAGGCGATAGAGCCGCTACGCAATATCTTTATGGGAGCCATCC 952
|||||
557 lyProIleSerProProGlnProSerValSerAlaThrPasnLys..P 573
|||||
953 CCATCAAGAGGATTGGAGCTGTCGGGGAAATACGGC..... 990'
|||||
573 roLeuThrSerPheGlySerAlaProSerSerGluGlyAlaLysAsnGly 589
|||||
991 .....TTGGCGCGCATCAGCGCACATCCTGTCAAGCG 1022
|||||
590 GlnGluSerGlyLeuGluIleGlyThrAspThrIle..... 601
1023 GTCCGAGATGGCGCGATCGCATTCGCCAAGAGGAATCCGCGTACAGG 1072
|||||
602 ...GlnPheGlyAlaProAlaSerAsnGlyAsnGluAsnGluValValP 617
1073 ACAATTTCCGATGGCGGATACGCGCAATACCCGTCCTCCCTTACCATTCC 1122
|||||
617 roValLeuSerGluLysSerAlaAspLysIleProGluPro..... 630
1123 CGAAATATCCGTTCAAACTTGGAGCAGCGCTTACGGCAAGAAACATCAC 1172
|||||
631 .....LysGluGlnArg..... 634
1173 CTCTCAACCGTGGCGCGCTCAACAGCGCAAAATGTCAAACCTGGCAGACC 1222
|||||
635 .....G 635
1223 AAGCCACCCGAGAGCAGCGGTACCGTTTGACGGTAAGGGTTTCCGAAT 1272
|||||
635 InLysGlnProArgAlaGly...ProIleLysAlaGlnLysLeuProAsp 650
1273 TTT.....GA 1277
|||
651 LeuSerProValGluAsnLysGluHisLysProGlyProIleGlyLysG 667
1278 GAAGCAGCTGAATATGATACGAGCTCGATTCGAATTAAGAAATATCGGGG 1327
|||||
667 uArgSerLeuLysAsnArgLysValLysAspAlaGlnGlnValGluProG 684
1328 GCGGTATACCTAAGGCTAAGCTGTGTTTGTATGCGAAGCCGAGATGGGAG 1377
|||||
684 LuGlyGlnLysProSerProAla..... 692
1378 GTTGATAGGAAGCTTAATAAATGACAACTCGTGACAGGTGGAGAAAAA 1427
|||||
693 ThrValArgSerThrAspProValThrLysGlu.....ThrLysAl 707
1428 TGTTCAGGAA.....ACGAGAAGAGGAGTCAGATA 1459
|||||
707 aValSerGluMetSerThrGluIleGlyThrMetIleSerValSerSera 724
1460 GTCAGTTTAAAGCCATGGCGCAACGAGAA.....TGGGAA 1494
|||||
724 laGluTyroGlyThrAsnAlaLysMetGluSerAlaArgLysAlaTrpGlu 740
1495 ATAATAACAGGTTAGATTATTAATCATTTATAGTGGTGTATATCAATAA 1544
|||||
741 AsnSerPro.....AsnValArgG 747
1545 GAAGGCACA 1554
747 uLysGlySer 750
```

seq\_name: p1r2:T30826

seq\_documentation\_block:

nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse

N:Alternate names: alpha-NAC protein

C:Species: Mus musculus (house mouse)

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 24-Nov-1999

C:Accession: T30826

R:Yotlov, W.V.; St-Arnaud, R.

Genes Dev. 10, 1763-1772, 1996  
A:Title: Differential splicing-in of a proline-rich exon converts alphanAC into a mus  
A:Reference number: Z20889; MUID:96312450  
A:Accession: T30826  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-2187 <YOT>  
A:Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1  
C:Genetics:  
A:Gene: Naca  
A:Map position: 10  
A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3  
A:Note: differential splicing converts alphanAC into a tissue-specific DNA-binding ac  
C:Keywords: alternative splicing; DNA binding; transcription factor

alignment\_scores:  
Quality: 141.50 Length: 569  
Ratio: 0.546 Gaps: 27  
Percent Similarity: 45.518 Percent Identity: 22.847

alignment\_block:  
US-09-303-518D-463 x T30826 ..

Align seg 1/1 to: T30826 from: 1 to: 2187

```
254 AAGCAATATCGGCTACATTTGTCGCTTTTCCGATCACGGGACAAATTC 303
|||||
833 LysAspThrSerAlaThrLeuSerLeuLysSerValProAlaValThrSe 849
304 CATTCGGCTTCGAAACCATGCTCATCTCGGATTTCTGACGAAGCCGG 353
|||||
849 rLeuSerProProLysAla.....Prov 857
354 TAGTCCCGTTGACGGATTCAGCTTTTACCAGTCCATTGGGACGGATACG 403
|||||
857 aAlaProSerAsnGluAla.....ThrIleVal 866
404 AACACCATCCGCGCGAGCTATGACGGCCACAGGGCGCGCTATCCC 453
|||||
867 ProThrGluIleProThrSerLeuLysAsnAlaLeuAlaAlaThrPr 883
454 GTCCTCCAAAGCGGAGGATATATACAGCTACGACATAAAAGCGGTGC 503
|||||
883 o.....LysGluThrLeuAlaThrSerIleP 892
504 CCAAAATATCGCTCAACCTGACCGCAACCCGACGACCGGACACGGC 553
|||||
892 roLysVal.....ThrSerProSerProGlnLysThr 902
554 TTGCGGACCGTTTCCACA.....ATCGCGCGCTATGCTGACGCAAGGA 597
|||||
903 ProLysSerValSerLeuLysGlyAlaProAlaMetThrSerLysLys... 918
598 GTAGCGCAGGATTCAACCGCGCACCCGATACAGCCCGAGCTGGACAG 647
|||||
919 ...AlaThrGluIleAlaAlaSerLysAspValSerProSerGlnPheP 934
648 ATCGGGCAATGCGG.....CCGAAGCCTTCAACGGGCACTG 682
|||||
934 roLysGluValProLeuLeuGlnHisValProProThrSerProLys 950
683 CAGATATCGTCAAAACATCATCGGCGCGGAGAGAAATTTGTCGGCGGA 732
|||||
951 SerProValSerAspThrLeuSerGlyAlaLeuThrSerProProLys 967
733 GCGGATCCCTGAGGATATAACCGAGGCTCAAACTTGTCTGTATGCA 782
|||||
967 sGlyProAlaThrLeu...AlaGluThrProThr..... 978
783 CGGCTTGGGCTCTTTCACCCGAAACAAAGATGGCGCGCATCAACGATT 832
|||||
979 .....TyrProLysLysSerProLysProAlaAla..... 988
```

```
833 TGGCAGATATGCGCAACTCAAGACTATGCGCGCAGCAGCATCCGCGAT 882
989 .....SerLysLysThrProAlaThrProSerProGl 999
883 TGGCGAGTCCAAACCCCAATG.....CCGCACACAGGCAT 917
999 uGlyValThrAlaValProLeuGluProProCysSerLysLysAlap 1016
918 AGAAGCGTCACCA..... 931
1016 roLysThrAlaAlaProLysGluSerSerAlaThrSerSerSerLysArg 1032
932 .....ATATCTTTTGGCAGGCATCCCATCAAGGGA..... 964
1033 AlaProLysThrAlaValSerLysGluLeuProSerLysGlyValThrAl 1049
965 .....TTGGAGCTGTCC..... 976
1049 aValProLeuGluLeuSerLeuProLeuLysGluThrSerLysSerAlat 1066
977 .....GGGAAATACGGCTGGCGGCATCAGCGCATCTGTCAAGC 1021
1066 hrProGlyGluLysSerAla.....SerSer 1074
1022 GGTCCGAGATGGCGCGCATCGCATTCGCGAAAGGAAATCCGCGCTCAGC 1071
1075 ProLysArgSerProLysThrAlaGlyProLysGluThrProProGlyGl 1091
1072 GACAAT.....TTGCCGATCGGCATAGCCAA 1100
1091 yValThrAlaValProProGluLeuSerLeuProProLysGluThrProG 1108
1101 ATACCGCTCCCTTACCATTCCGGAATATCGGTCAAACTTGGAGCAGC 1150
1108 InAsn.....AlaThrProAsnGluSerLeuAlaAlaSerSerGln 1121
1151 GTTACGGCAAGAAACATCACTCTCTCAACC.GTGCCTCGCTCAACGG 1199
1122 LysArgSerProLysThrSerValProLysGluThrProProGlyGlyVa 1138
1200 CAAAAATGTCAACTG.....GCAGACCAACGCCACCGAAGA 1237
1138 lThrAlaMetProLeuGluLeuProSerAlaProGlnLysAlaProLysT 1155
1238 CAGGCGTACCGTTTACGGTAAAGGTTTCCGAATTTGAGAAGCAGCTG 1287
1155 hrAlaValPro.....LysGlnIleProThrProGluAspAlaVal 1168
1288 AAATAT.....GATACGAAGCTCGATATTCAAGAAATTATCGGGGCGG 1331
1169 ThrIleLeuAlaGlySerProLeuSerProLysLysAlaSerLysThrAl 1185
1332 TATACCTAAGCTAAGCCTGTGTTTATGCGCAACCGAGATGGAGGTG 1381
1185 aAlaProLysGluAlaProAlaThrProSerValGlyValIleAlaValS 1202
1382 ATAGGAAGCTTAAT.....AAATTGACAACTGCTGACGAGGTGGAG 1422
1202 erGlyGluIleSerProSerProLysLysThrSerLysThrAlaAlaPro 1218
1423 AAAAATGTTTCAGAACAGAGAGAGAGTCTAGACTAGTCTTAAGC 1472
1219 LysGluAsnSerAlaThrLeuProProLysArgSerProLysThrAlaAl 1235
1473 CCATCGCAACAGAGAATGGGAAATATAACAGGGTTAGATTTTATCAT 1522
1235 aProLysGluThrProAlaThrSerSerGluGlyValThr.....A 1249
1523 TTATAGGTGGTGATATCAAT.....AAGAAAGC 1551
1249 laValProSerGluLeuSerProSerProProThrProAlaSerLysGly 1265
```

```
1552 .....ACAGTAACA.....GGAGGGCATAGTCTAAACCGTGG 1583
1266 ValProValThrLeuThrProLysGlyAlaProAsnAlaLeuAlaGluSe 1282
1584 TGATGTA.....CGGGTGATACAAACACCTCGGCACCTGATAAACATG 1627
1282 rProAlaSerProLysLysValProLysThrAlaAlaProGluGluThrs 1299
1628 GGGTTTATCAAGCAGCAGTGGAAATTAAGCCTGATGGAAGTTGGGAG 1677
1299 erThrThrProSerProGlnLysIleProLysValAlaGlyProLysGlu 1315
1678 GTGAAACGAAAAAGGTGGAAAGTATGACCAAGCACACCATGTGCC 1727
1316 AlaSerAlaThrProProSerLys.....LysThrProLysThrAlaValPr 1331
1728 AAAAGAT 1734
1331 oLysGlu 1333
seq_name: pir2:S35047
seq_documentation_block:
mucin JUL7 - human
C:Species: Homo sapiens (man)
C>Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C:Accession: S35047
R:Dufosse, J.; Porchet, N.; Audie, J.P.; Guyonnet Duperat, V.; Laine, A.; van-Seuning
Blochem. J. 293, 329-337, 1993
A:Title: Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic altern
A:Reference number: S35047; MUID:93343858
A:Accession: S35047
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-543 <DUE>
A:Cross-references: EMBL:X74370; NID:g407081; PIDN:CAA52408.1; PID:g407082
A>Note: the authors translated the codon AAA for residue 63 as Gln and CGG for residu
A>Note: the amino acid sequence from Fig. 2 is inconsistent with the nucleotide seque
alignment_scores:
Quality: 141.00 Length: 467
Ratio: 0.698 Gaps: 27
Percent Similarity: 43.255 Percent Identity: 24.625
alignment_block:
US-09-303-518D-463 x S35047 ..
Align seg 1/1 to: S35047 from: 1 to: 543
68 ACGCTCAGATTGGCAACGATCCCTTTATCCGGCAGGTTCTCGACCGT 117
138 ThrProGluThrThrHisThr.....SerThrVa 147
118 CAGCATTTTGAACCCGACGCGAAATACCACTATTTCGGCAGCAGGGGGA 167
147 lLeuThrThrLysAlaThr.....ThrThrArgAlaThrSerSerThrS 162
168 GCTTGCNAGCGCAACGCCATATCGGATTGGGAACATACAAAGCCATC 217
162 erThrProSerSerThrProGlyThrThrTrpIleLeuThrGluLeuThr 178
218 AGTTGGCCACCTGATGATTCACAGCGGCGCG..... 250
179 ThrAlaAlaThrThrThrAlaGlyThrGlyProThrAlaThrProSerSe 195
251 .....TTGAAGAAATATCGGTACATTTGCCG 278
195 rThrProGlyThrThrTrpIleLeuThrGluLeuThrThrThrAlaThrT 212
279 CTTTTCGATCAGGGG.....ACAAATTCATTCGCCCTTCGACA 319
212 hrThrAlaSerThrGlySerThrAlaThrLeuSerSerThrProGlyThr 228
```



```
95 pGlnAspGlyAlaProArgArgHisLeuArg.....ArgA 109
    ::::: ||||| ::::: |||
445 GGCTATCCCGCTCCCAAGGCGGAGGATATATACAGCTACGACAT... 491
    ||::: ::::: |||||
109 rgValGlyGlyHisArgGlyArgAsn.....ArgHisAla 120
    ::::: ::::: |||||
492 ...AAAAGCGGTGCCCAAAATATCCGCTCAACCTGACCGCACACCGCA 538
    ||||| ::::: |||
121 GlyAspArgArgAlaProGlyValAspSerArgLeuArgGlnHisG 137
    ||||| ::::: |||
539 GCACCGGACACAGCGCTTCCCGACCGTTCCACAATGCCGGCGCTATG 588
    ||||| ::::: |||
137 nHis.....ProArgGlyArgHisAlaSerA 146-
589 AC.....GCAAGGAGTAGCGGACGCGATCAAA 614
    ||
146 spArgValGlnAspGlyAlaHisProArgArgGlnArgLeuArgGlu 162
615 ACGGCGCACCGATACAGCCCGGAGCTGGACAG.....ATCGGGCA 655
    ||||| ::::: |||
163 ProArgHisAlaGlyArgProArgArgGlnProProArgArgGlyAr 179
    ||||| ::::: |||
656 ATGCGCGGAAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATC 705
    : ||| ::::: |||||
179 gSerArgGlyThrHisArgArgHisLeuArgGlnAlaProArgProAla 196
706 GCGCGCGGAGAGAAATGTGCGCGGAGCGGATCGCGTGCAGGGTATAAG 755
    ||||| ::::: |||||
196 alArgGlyProAspGluAspGlnAlaArgGluPheArgGlyProArgHis 212
756 CGAAGGCTCAACATTCGTGTCATGACAGG.....CTTGGGTCTGC 796
    ||||| ::::: |||||
213 ArgArgGluArgHisProProThrAlaArgAspValLeuArgGlyGlu 229
797 TTTCCACGGAACAAAGATGCGCGCATCAACGATTGGCAGATATGGCG 846
    ||| ||||| |||
229 oGlyHis.....GlyAspGlyHisHis..... 236
847 CAACTCAAGACTATGCCGAGCAGCATCCGGATTTGGCAGTCCAAAA 896
    ::::: ||| ||| |||||
237 .....LeuGlyArgArgGlyArgProArg 245
897 CCCAATGCCGACAGGCATAGAC.....CGTCA 928
    ||||| ||| |||||
246 ProGlnGlyArgGluAlaGlyArgGlyAlaHisProProGlnValArgA 262
929 GCAATATCTTTATGGCAGCATCCCATCAAGGGATT..... 966
    ::::: ||||| ::::: |||||
262 laArgIleTyrLeuAlaAlaGlyGluAlaArgGlyLeuProGluProArg 278
967 .....GGAGCTGTCCGGGAAATA 986
279 ProLeuGlyValArgThrValHisArgGlyArgLeuArgGlyArgVa 295
987 CGGCTTTGGGCGGATCAGCGCACATCTGTCAAGCGGTGCGACATGGCG 1036
    ||| |||||
295 lGlyGlnAlaGly.....ProArgProGlnValProG 306
1037 CGATCGCATTCGGAAGGGAATCCGCGTCAGCGACAATTTGCGCGAT 1086
    : ||||| ::::: |||
306 lyAspPheAlaProGlnGlyGlu.....Asp 314
1087 GCGGCATACGCCAAATACCGTCCCTTACCATTCCGAAATATCGGTTTC 1136
    ::::: ||| |||||
315 SerGluArgGluThrProProArgProHisSerArgLysArgArgAs 331
1137 AAATTGTGAGCAGCGTTACGGCAAGAAACATCATCTCTCTCAACCGTGC 1186
    ::::: |||||
331 pThr...GlyAla.....HisHisArgHisTrpArgAr 341
1187 CGCGGTCAACGCGCAAAATGTCAACTGGCAGA.....COAAGCGCAC 1230
    ::::: ||| ::::: |||
```

```
341 gArgArgArgValArgHisArgGluGlyAlaLeuProAlaAlaHisP 358
1231 CCGAAGACAGCGCT 1244
    ||::: |||||
358 roAspAspArgArg 362
seq_name: pir2:A46299
seq_documentation_block:
tyrosine kinase substrate disabled - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 22-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 01-Aug-1997
C:Accession: A46299
R:Gertler, F.B.; Hill, K.K.; Clark, M.J.; Hoffmann, F.M.
Genes Dev. 7, 441-453, 1993
A:Title: Dosage-sensitive modifiers of Drosophila abl tyrosine kinase function: prosp
A:Reference number: A46299; MUID:93194063
A:Accession: A46299
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-2411 <GER>
A:Cross-references: GB:L08845
A>Note: sequence extracted from NCBI backbone (NCBIP:127163)
C:Keywords: alternative splicing; phosphoprotein
alignment_scores:
    Quality: 138.50      Length: 558
    Ratio: 0.523        Gaps: 38
    Percent Similarity: 47.491      Percent Identity: 23.835
alignment_block:
US-09-303-518d-463 x A46299 ..
Align seg 1/1 to: A46299 from: 1 to: 2411
266 GCTACATTGTCCGCTTTTCCGATCAGGGCACAAATTCATTCGC...CC 312
    ||||| ||||| ||| ||||| ::::: ||
1405 AlaThrSerSerAlaAlaProSerArgGlyAlaProGlyLeuHisThrPr 1421
313 TTTCG.....ACAACCATGCCTCACATCCGATTCTTCGAGA 347
    |||| ::::: |||| ||||| |||
1421 oSerGlnPheAsnAspValSerThrSerPro...ilePro...LeuGlnL 1436
348 AGCCG.....GTAGTCCCGTTGACGGATTACGCCCTTTACC 382
    |||| ::::: |||||
1436 ysProGlyMetGlyProSerProValProSerGlnLeuSerAlaValSer 1452
383 GCATCCATTGGCAGGATACGAACACCATCCCGCGGCGGCTATGACGG 432
    ||| ||| |||||
1453 GlnLeuIleAspThrAlaThrLys.....GlnMetMetGl 1464
433 CCACAGGCG.....CGGCTATCCCGCTCCCAAGG 464
    | ::::: |||||
1464 yAspLysAspArgGluLysGlnSerTrpAlaThrPheAspSerProLysA 1481
465 CGCAGGAGTATATACAGCTACGACATAAAGCGTTGCCCAAAATATCC 514
    ||::: |||||
1481 laLysGlyLysAlaArgLeuThrLeuProProProProProAlaSer 1497
515 GCCTCACTGACCGACAAACCGCACACCGGACACCGCTTTCGCG... 559
    ::::: ||| |||||
1498 AsnThrSerGlnProAspThr.....GluAsnArgLeuAlaValAr 1511
560 .....ACCGTTCCCAATCCGCGCTATGTCGACGCAAGG 596
    ||||| ::::: |||
1511 gileProGlyMetThrAlaGlyGlnSerAsnSerValValGlyArgArgA 1528
597 AGTAGGCGAGGATTAAGCGCGCCACCGCATACAGCCCGGAGCTGACA 646
    :: ||||| |||||
1528 rgSerAlaThrThrSerSerSerArgAspLeuSerPro...TrpAsp 1543
647 GATCGGCAATGCCCGCAAGCGCTTCAACGGCACTGCAGATATCGTCAA 696
```



2006年6月



C:Accession: T15316

R:Du, Z.

submitted to the EMBL Data Library, November 1995

A:Description: The sequence of C. elegans cosmid B0302.

A:Reference number: Z18328

A:Accession: T15316

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1223 &lt;DUZ&gt;

A:Cross-references: EMBL:U41032; PID:g1086728; PIDN:AAA82367.1; CESP:B0302

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:B0302.1

A:Introns: 50/3; 106/3; 153/3; 180/3; 243/3; 399/2; 438/3; 482/1; 520/2; 545/1; 589/1; 6

alignment\_scores:

Quality:	137.50	Length:	523
Ratio:	0.593	Caps:	24
Percent Similarity:	44.359	Percent Identity:	22.753

alignment\_block:

US-09-303-518D-463 x T15316 ..

Align seg 1/1 to: T15316 from: 1 to: 1223

```
12 CGCGAAATATCCCTTATCTGTCCATATCGCAGTGTGCTGCGGATGC 61
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
715 ProAlaAsnIleProCysLeuValProThrProAlaProAlaProAl 731
62 ATGCACACGCTTCAGATTGGCAACGATCCCTTATCCGCGAGGTTCTC 111
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
731 a..HisPheSerGln.....ProValSerSerGlnArgVa 742
112 GACCGTCAGCATTTCCGAAC..... 130
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
742 lalaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 759
131 .....CGACGGGAAATACACCTATTCGCGCAGCA 160
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
759 ysGlyAsnLeuAsnLysArgProThrGlyThrAlaProProSerAsn 775
161 GGGGGAGCTTCCNAGCGCAACGCCATATCGGATTCGGGAACATACAA 210
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
776 GlyPheAsnAlaProArgAlaAspValAlaPro..... 786
211 AGCCATCAGTTGGGCGCACCTGATGATTCACAGCGCGCGTTGAAGGAA 260
787 .....ValGlnGlnArgProIleSerSera 795
261 TATCG...GCTACATGTGCGCTTTTCGATCAGGGGCACAAATTC... 304
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
795 laSerIleProAlaLeuGlnProGlnProIleGlnHisIleGlnLysPro 811
305 ATTCGCCCTTCGACAAACCATCCCTCAGATTCGGATTCGACGAAGCGGT 354
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
812 IleGlnProGlnGlnVal.....ArgIleProProSerThrAlaProVa 826
355 AGTCCCGTTGAGGATTTCAGCTTTTACCGCATCCATTGGGACGATACGA 404
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
826 l.....GlnLysProValGlnValSerAlaProThrHisSera 839
405 ACACATCCCGCGGACGCTATGACGGGCCACAGGGGGCGGCTATCCCG 454
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
839 snValAlaProThrThrSerSerGlnAlaSerAlaAspAlaArgAsnPro 855
455 CTCCTCAAGGCGGAGGATATATACAGCTACGACATAAAGGCGTTGCC 504
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
856 leuProPro.....LysThrSerPr 862
505 CAAATATCCGCCTCAACCTGACCGGACACCGCAGCCGCGACACCGCT 554
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
862 oProValSerAsnThr.....ProIleThrValAlaProValHisAlaA 877
```

```
555 TGCGACCGCTTCCACAATCCGCGCGCTATGCTGACGCAAGGAGTAGCG 604
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
877 laProThrThrSer..... 881
605 ACGGATTCAACACGCGCCACCGGATACAGCCCGAGCTGGACACATCGGC 654
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
882 .....AlaProSerThrSerValValTh 889
655 AATGCCCGCGAAGCTTCAACGACCTGCAGATATCGTCAAAAACATCAT 704
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
889 rArgArgProThrSerThrAla...GlnMetSerAspGluGluArgA 905
705 CGCGCGCGGAGGAGAAATTCGCGCGGAGGATCGCGTCGAGGATATAA 754
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
905 rgSerArgIleAlaMetAspIleSerSerAlaLeuPro..... 917
755 GCGNAGCTCAACATTCGTGTCATGACGCGGCTTGGTCTGCTTTCAC 804
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
918 AlaProSerAlaLeuLeuTyGlySerAsnSerThrSerSerLeuProSe 934
805 GAAACCAAGATGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCA 853
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
934 rAlaAlaValSerThrAlaSerSerValProSerThrAlaArgAspAsn 951
854 .....AAGACTATCCGCGACGACATCCGCGATTTGGCAGTCCAAACC 898
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
951 roValGluThrArgProSerGlnProHisVal.....Thr 962
899 CCAATGCCGACAGGATAGAACGCGTCGACGAATATCT...TTATGGCA 945
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
963 MetProProLysLysSerSerGluProIleLeuSerSerGluValLeuG 979
946 GCCATCCCATCAAGGATTTGGAGCTGTCGCGGGAATAACGCGTTGG 995
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
979 nProThrArgLeuProSerSerAlaThrThrSerGlnAlaLysProValThrG 996
996 CGCATCACGCGCATCTCTGTCAGCGGTCGCGAGATGGCGCGCATCGCAT 1045
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
996 lnProIleArgHis..... 1000
1046 TGCGAAAGGAATCCGCGCTCAGCGACAAATTTTCCCGAT..CGGCATA 1094
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1001 .....ProSerProValAlaThrValIleProThrAlaValVa 1014
1095 CGCAATACCCGCTCCCTTACCATTCGCGAAATATCCGTTCAAACTG 1144
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1014 lasPlyLysProVal.....SerGlnAsnGlnGlySerAsnValP 1028
1145 AGCAGGTTACGGCAAGAAACATCACCTCTCTCAACC.....GTGCG 1188
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1028 ro.....LeuPheAsnIleThrAsnSerSerAsnGlyTyPro 1040
1189 CCGTCAACGGCAAAATGTCAAACTGGCAGACCAACGCCACCCGAGAC 1238
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1041 GlnLeuAsnGly.....TyProAsnTy 1048
1239 AGCGGTACCGTTTGACGGT.....AAAGGT 1264
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1048 rGlyAsnGlyPheGlnAlaTyGlyTyGlyMetAsnTyHisGlnGlyT 1065
1265 TTCCGAATTTTGAAGACGCTGAATATGATACGAAGCTCGATTTCAA 1314
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1065 yrProGlyTyGlnGlyTyAsnSerTyGlyAsnGlyMetGlyGlnLeu 1081
1315 GAATATCGGGGGGGTATACCTAGGCTAAGCTGTGTTTGTATCGGAA 1364
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1082 AlaLeuThrHisAsnAlaValThrSerLeuProProLeuValProSerg 1098
1365 ACCGAGATGGAGGTGTAGAGAGCTTAATAAATGACAACTCGTGAGC 1414
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1098 uAsnArgPheSerGly.....ThrAlaGlnProL 1108
```

1415 AGGTGGAGAAAATGTCAGGAA.....ACGAGAAGAGGAGTCTAG 1455  
|||||.....  
1108 euGlyGluSerAspIleMetGluPheLeuGlyThrGlnGlnArgGlnAla 1124  
|||||.....  
1456 AGTAGTCAGTTTAAAGCC 1473  
|||||.....  
1125 GlySerSerSerArgAla 1130  
|||||.....

seq\_name: pir2:S51342

seq\_documentation\_block:

verprolin - yeast (*Saccharomyces cerevisiae*)  
N:Alternates names: prolin-rich protein VRP1; protein L8300.13; protein YLR337c  
C:Species: *Saccharomyces cerevisiae*  
C>Date: 23-Feb-1995 #sequence\_revision 11-Aug-1995 #text\_change 23-Mar-2001  
C:Accession: S51342; S39626; S57435  
R:Du, Z.  
submitted to the EMBL Data Library, December 1994  
A:Description: The sequence of *S. cerevisiae* cosmid 8300.  
A:Reference number: S51339  
A:Accession: S51342  
A:Molecule type: DNA  
A:Residues: 1-817 <DUZ>  
A:Cross-references: EMBL:U19028; NID:g609380; PID:g609392; MIPS:YLR337c  
R:Donnelly, S.F.H.; Pocklington, M.J.; Pallotta, D.; Orr, E.  
Mol. Microbiol. 10, 585-596, 1993  
A:Title: A proline-rich protein, verprolin, involved in cytoskeletal organization and cell division in *S. cerevisiae*  
A:Reference number: S39626; MUID:95058201  
A:Accession: S39626  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-307, 'R', 309-349, 'R', 351-688, 'E', 690-709, 'HLRWIPVPLIAPVKLTNNGYFLQVDRRCNTSII'  
A:Cross-references: EMBL:226645; NID:g414785; PID:CAA81388.1; PID:g439289  
R:Munn, A.L.; Stevenson, B.J.; Geli, M.I.; Riezman, H.  
submitted to the EMBL Data Library, June 1995  
A:Description: ends5, ends6, and ends7: mutations that cause actin delocalization and block cell growth in *S. cerevisiae*  
A:Reference number: S57435  
A:Accession: S57435  
A:Molecule type: DNA  
A:Residues: 1-162, 'F', 164-817 <MUN>  
A:Cross-references: EMBL:X87806; NID:g871534; PID:g871535  
A:Experimental source: strain W303  
C:Genetics:

A:Gene: SGD:VRP1; MDP5; END5  
A:Cross-references: SGD:S0004329; MIPS:YLR337c

A:Map position: 12R

F:5-14/Region: proline-rich

F:77-85/Region: proline-rich

F:114-180/Region: proline-rich

F:216-245/Region: proline-rich

F:305-336/Region: proline-rich

F:349-357/Region: proline-rich

F:372-382/Region: proline-rich

F:396-406/Region: proline-rich

F:421-445/Region: proline-rich

F:518-528/Region: proline-rich

F:567-577/Region: proline-rich

F:608-621/Region: proline-rich

F:649-661/Region: proline-rich

F:678-685/Region: proline-rich

F:704-710/Region: proline-rich

alignment\_scores:

Quality:	137.00	Length:	372
Ratio:	0.878	Gaps:	18
Percent Similarity:	41.935	Percent Identity:	23.656

alignment\_block:

US-09-303-518d-463 x S51342 ..

Align seg 1/1 to: S51342 from: 1 to: 817

275 TCCGCTTTTCCGATCACGGGCACAAATTCATTCG...CCTTCGACAAC 321  
|||||.....  
64 SerAlaSerGlySerGlyThrValSerSerGlyProSerMetSe 80  
|||||.....  
322 ATGCTCCTCACATTCGATTCGACGAAGCCGGTA..... 355  
|||||.....  
80 rAlaProProlleProGlyMetGlyAlaProGlnLeuGlyAspIleLeuA 97  
|||||.....  
356 .....GTCCCGTTGACGGATTTCAGCCTTTACCGCATCCATTTGGGACG 397  
|||||.....  
97 laglyGlyIleProLys.....LeuLysHisIleAsnAsn 108  
|||||.....  
398 GATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGCGC 447  
|||||.....  
109 AsnAlaSerThrLysProSerProSer.....AlaSerAl 120  
|||||.....  
448 TATCCCGCTCCCAAGGGCGGAGGATATATACAGCTACGACATAAAAGG 497  
|||||.....  
120 aProProllePro.....GlyA 126  
|||||.....  
498 CGTTCGCCCAAAATATCCGCTCAACCTGACCGACAAACCGCAGCAGCGAC 547  
|||||.....  
126 laValProSerValAlaAlaPro.....ProlleProAsnAlaProLeu 140  
|||||.....  
548 AACGGCTTCGCGACCGTTTCCACAAATGCGCGCTATGCTGACGCAAGA 597  
|||||.....  
141 SerProAlaProAlaValProSerIleProSer..... 151  
|||||.....  
598 GTAGGCGGCGGATTCAACGGCGCCACCGATACAGCCCGAGCTGACAGAC 647  
|||||.....  
152 .....SerSerAlaProPro.....IleProAspIleProS 162  
|||||.....  
648 ATCGGCAATGCGCGCG.....AAGCCTTCAACGGCA 679  
|||||.....  
162 erSerAlaAlaProProlleProlleValProSerSerProAlaProPro 178  
|||||.....  
680 CTGCAGATATCTCAAAAACATCA..... 703  
|||||.....  
179 LeuProLeuSerGlyAlaSerAlaProLysValProGlnAsnArgProHi 195  
|||||.....  
704 .....TCGGCGCGGAGGAGAAATTCGCG 728  
|||||.....  
195 sMetProSerValArgProAlaHisArgSerHisGlnArgLysSerSerA 212  
|||||.....  
729 CGCAGCGGATGCGTGCAGGTATAAGCGAAGGCTCAAAACATTGCTGTCA 778  
|||||.....  
212 snIleSerLeuProSer..... 217  
|||||.....  
779 TGCACGGCTTGGGTCGTGCTTTCACCGGAAACAAAGATGCGCGCATCAAC 828  
|||||.....  
218 .....ValSerAlaProProLeuPro..... 224  
|||||.....  
829 GATTTCGCAGATATGCGCAACTCAAAAGACTATGCGCGACAGCAGCATCCG 878  
|||||.....  
225 .....SerAlaSerLeuProThrHisValSerA 234  
|||||.....  
879 CGATTGGCGAGTCCAAAACCCCAATGCCG..... 907  
|||||.....  
234 snProGlnAlaProProProProProProThrProThrIleGlyLeuAsp 250  
|||||.....  
908 CACAAGGCATAGAACGGCGTCAGCAATATCTTTATGGCAGCCATCC..... 952  
|||||.....  
251 SerLysAsnIleLysProThrAspAsnAlaValSerProProSerSergI 267  
|||||.....  
953 .....CCATCAAGGGGATTGGAGCTGTCGGGGGAAATACGGCTTGGCGCG 998  
|||||.....  
267 uValProAlaGlyGlyLeuProPheLeuAlaGluIleAsnAlaArgArgS 284  
|||||.....  
999 CATCAGCGCACATC.....CTGTCAAGCGGTGCGCAGATGGCGCGCA 1039  
|||||.....  
284 erGluArgGlyAlaValGluGlyValSerSerThrLysIleGlnThrGlu 300  
|||||.....  
1040 TCGCATTTGCCGAAAGGAAATCCGCG.....TCAGCGACAATTTT 1080









n seg 1/1 CO: E80185 FROM: I CO: 492

86 AGCATCCCTTTATTCGGCAGGTTTCTCGACCGTCAGCAATTTCGAACCCGAC 135  
||| |||:  
68 ThrAlaGluSerSerSerThrLysSerAlaThrMetSerGlySerThr 84  
136 GGGAAATACCACCTATTTCGCAGCAGGGGGGACTTGCNMGGCCAACGG 185  
||||| |||:  
84 ThrHisThrThrSerSerAlaThrAlaSerSerThrAlaSerThrSert 101  
||| |||:  
186 CCATATCCGANTGGCAACATACAAGCCATCAGTTGGCGCACCTGATCA 235  
:: ||| ::|||:  
101 hrSerSerTySerThrSerTySerThrSerSerThrLysThr..... 115  
236 TTCAAAGGGCGCGTTGAAGGAATAATCGGTACATTCGCGCTTTTCOC 285  
:: ||| |||:  
116 ..ThrThrMetThrGlySerThrIleSerThrThrAlaSerAlaAlaPr 131  
286 GATCAGC.....\*..GCAC 296  
||| :  
131 othrSerThrAlaSerThrSerThrSerSerTySerThrSertYrSert 148  
297 CAAATTCATTCGGCCTTCGACACCAACATCGCTCACATTCGATTCGACG 346  
|||:  
148 hrSerSerThrLysThrThrValThrThrValThrGlySerThrIleGlyThrThr 164  
347 AAGCCGGTAGTCGCTTGAGGATTCAGCCTTTACCGCATCCATCCGAG 396  
:: ||| |||:  
165 AlaSerAlaAlaProThrSerThrSerThrSerThrAlaAsnSerSeral 181  
397 GGATACGACACCATCCCG.....CCGACGCTATCAGCG 431  
:: |||:  
181 aserSerThrThrAsnProSerSerGlySerLysProThrAlaMetThrG 198  
432 GCCACGGCGCGCGCTATCCGCTCCCAAGCGCGAGGATATATACA 481  
||| |||:  
198 lyThrThrAlaAsnThrSerProSerAlaProThrSerSerProSerThr 214  
482 GTACGACATAAAAGCGGTGCCAANAATATCCGCTCAACCTGACCGAC 531  
|||:  
215 ThrAsnSerSerSerThrAlaAlaTyThrSerSerGlySerLysProTh 231  
532 AACCCGACACCGGACACGCTTCCGACCGCTTCCACAAATCCCGCGC 581  
|||:  
231 rThrValThr.....ArgThrThrAlaAsnThrSerSerSera 244  
582 TATGTCGACGAAGAGGTAGCGACGGATCAACAGCGCCACCGCATACA 631  
::|||:  
244 la.....SerThrSerSerAlaSerProThrAsnSer 254  
632 GCCCGCAGGTGG...ACAGATCGGGCAATGCGCGGAGAGCTTCAACGGC 678  
|||:  
255 SerThrSerThrProThrAsnSerSerAlaGlySerLysProThrThrMe 271  
679 ACTGCAGATATCGTCAAAACATCATCG.....G 707  
271 thrGlyThrThrThrAsnThrSerSerThrThrThrThrSerSerAlas 288  
708 CGCGCGAGGAGAAATTGCGCGCAGCGAGTTCGCGTACGGGTATAAGCG 757  
|||:  
288 erThrThrLysSerSerSerSerAlaThrAsnSerSerSerGlySer 304  
758 AAGGCTCAACATCTCTGTCATGACGCGCTTGGGTCTGCTTCCACCGAA 807  
||| |||:  
305 LysProSerThrLeuSerThrThrThrAlaTyThrAlaThrThrSerSe 321  
808 ACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCACTCAAGA 857  
|||:  
321 rProThrAlaGluProSerThrThrThrAlaSerLysProAlaThrSers 338  
858 CTATGCCGCGAGCGCATCCCGGATTTGGCAGTCCAAACACCCCAATGCCG 907  
||| |||:  
338 erThrProProAlaPro.ProThrIleIleValSerLysProArg...Ar 353

403 rgHisArgSerSerArgAspGlyTyrGluGlnSerArgTyrSerGln 419







C:Species: Escherichia coli  
C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 08-Dec-2000  
C:Accession: I54632  
R:Providence, D.L.; Curtiss, R.  
A:Title: Immun. 62, 1369-1380, 1994  
A:Title: Isolation and characterization of a gene involved in hemagglutination by an avian influenza virus  
A:Reference number: I54632; MUID:94178945  
A:Accession: I54632  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-1377 <RES>  
A:Cross-references: GB:I27423; NID:9469235; PIDN:AAA24698.1; PID:9469236  
C:Superfamily: IgA-specific metalloendopeptidase

alignment\_scores:  
Quality: 134.00 Length: 707  
Ratio: 0.428 Gaps: 36  
Percent Similarity: 44.272 Percent Identity: 20.509  
alignment\_block:  
US-09-303-518d-463 x I54632 ..  
Align seg 1/1 to: I54632 from: 1 to: 1377  
109 CTCGACGCTCAGCATTCGAA.....CCCGACGGGAATA 143  
|||||.....  
182 LeuAspLysGluArgTyrProValPheTyrArgLeuGlySerGlyThrG1 198  
144 CCACCTATTCGGCAGCAGGGGAGCTTCCNAGCGCAACGCCATATCG 193  
.....  
198 nTyrIleLysAspSerAsnGlyGlnLeuThrGlnMetGlyAlaTyrS 215  
194 GA.....TTGGAAACATACAAAGCCATCAGTTGGC 225  
:::.....  
215 eTrpLeuThrGlyGlyThrValGlySerLeuSerTyrGlnAsnGly 231  
226 CACCTGATG..... 234  
232 GluMetIleSerThrSerSerGlyLeuValPheAspTyrLysLeuAsnG1 248  
234 ..... 234  
248 yAlaMetProIleTyrGlyGluAlaGlyAspSerGlySerProLeuPheA 265  
235 .....ATTCAACAG..... 243  
265 laPheAspThrValGlnAsnLysTrpValLeuValGlyValLeuThrAla 281  
244 .....GCGGCGCTTGAAGGAATATCGGCTACATTTGTCGCTTTTC 284  
|||:::.....  
282 GlyAsnGlyAlaGlyGlyArgGlyAsnAsnTrpAlaValIleProLeuAs 298  
285 CGATCAGCGGCAAAATTCATTCGCTTCCGCTTCCGCAACCATGCC..... 327  
: |||||.....  
298 pPheIleGlyGlnLysPheAsnGlu.....AspAsnAspAlaProValT 313  
328 .....TCACATTCGATTC 342  
313 hrPheArgThrSerGluGlyAlaLeuGluTrpSerPheAsnSerSer 329  
343 GACGAAGCGGTAGTCCCGCTTACCGGATTCAGCCCTTTACCGCATCCATG 392  
|||||:::.....  
330 ThrGlyAlaGlyAlaLeuThrGlnGlyThrThrThrThrAlaMetHis... 345  
393 GGACGGATACGAACACCATTCCTCCCGCGCGGCTATGACGGGCCACAGGCG 442  
346 .....GlyGlnGlnGlyA 350  
443 GCGGCTATCCCGCTCCCAAGCGCGGAGGATATATACAGCTACGACATA 492  
::: |||.....  
350 snAspLeuAsnAlaGlyLysAsnLeu.....IlePheGln..... 361

493 AAGGCGTTGCCCAAAATATCGCTCAACCTGACGCAACACCGCAGCAGC 542  
:::|||||.....  
362 .....GlyGlnAsnGlyGlnIleAsnLeuLysAspSerValSerG1 375  
543 CGGACAACAGCTTGGCGACCGCTTCCACAAATCCCGCGCTATGCTGACGC 592  
||| |||||.....  
375 nGly.....AlaGlySerLeuThrPheA 383  
593 AAGGAGTAGCGCAGGATTCACACGCGCCACCGGATACAGCCCCGAGCTG 642  
::: |||||.....  
383 rg.....AspAsnTyrThrValThrThr..... 390  
643 GACAGATCGGCAATCCCGCAAGCTTCAACGCGCACTGCAGATATCCT 692  
::: |||||.....  
391 .....SerAsnGlySerThrThrThrGlyAlaGlyIleValVa 403  
693 CAAAAACATCATCGCGCGGAGGAGAAATTTGTCGGCGCA...GGCGATG 739  
I ||| |||||.....  
403 lAspAsnGlyValSerValAsnTrpGlnValAsnGlyValLysGlyAspA 420  
740 CCGTGCAGGTATAGCGAGGCTCAACATTCCTGCTCATGCACGCGCTTG 789  
::: |||||.....  
420 snLeuHisLysIleGlyGlyThr.....LeuThrValGlnGlyThr 434  
790 GGTCTG.....CTTCCACCGAAACACAGATGCGCGCATCAA 827  
|||::: |||||.....  
435 GlyIleAsnGluGlyGlyLeuLysValGlyAspGlyValValLeuAs 451  
828 CGATTTGGCAGAT...ATGCGCACTCAAGACTATGCGCGCAGCAGCCA 874  
I ||| |||||.....  
451 nGlnGlnAlaAspAsnLysGlyGlnValGlnAlaPheSerSerValAsnI 468  
875 TCCGCGATTCGGCAGTCCAAACCCCAATGCCGCAACAGCATAGAGCC 924  
||| |||||.....  
468 le.....AlaSerGlyArgProThrValValLeuThrAspGluArg 481  
925 GTCAGCAATATCTTTATGCGAGCATCCCATCAAGAGGATGGAGCTGT 974  
||| |||||.....  
482 GlnValAsnProAspThrValSerTrpGlyTyrArgGly..... 494  
975 CCGGGGAAATACGGCTTGGCGGC.....ATCAGGCGACATCTCTGCA 1018  
||| |||||.....  
495 ....GlyThrLeuAspValAsnGlyAsnSerLeuThrPheHisGlnLeuL 510  
1019 AGCGGTGCGCAGATGCGCGCATC..... 1041  
510 yAlaAlaAspTyrGlyAlaValLeuAlaAsnValAspLysArgAla 526  
1042 .....GCATTGCGGAAAGGAAATCCCGCTGACG... 1071  
527 ThrIleThrLeuAspTyrAlaLeuArgAlaAspLysValAlaLeuAsnG1 543  
1072 .....GACATTTTCCGATGCGGATACGCGCAATACC 1105  
543 yTrpSerGluSerGlyLysGlyThrAlaGlyAsnLeuTyr...LysTyrA 559  
1106 CGTCCCTTACCATTCGCCGAAATATCGTTCAAACTTGGAGCAG...CGT 1152  
559 snAsnProTyrThrAsnThrThrAspTyrPheIleLeuLysGlnSerThr 575  
1153 TAGGCG.....AAGAAACATCATCCCTCTCAACCGT 1184  
576 TyrGlyTyrPheProThrAspGlnSerSerAsnAlaThrTrpGluPheVa 592  
1185 GCGCGCTCAACCGGCAAAATGTCAAACTGGCAGACCAACGCCACCCGA 1234  
592 lGlyHisSerGlnGlyAspAlaGlnLysLeuValAlaAspArgPheAsnT 609  
1235 AGACAGCGGTACCGTTTGGCGGT.....AAGGGTTTCCGAAATTTTGG 1278  
609 hrAlaGlyTyrLeuPheHisGlyGlnLeuLysGlyAsnLeuAsnValAsp 625  
1279 AAGCAGCTGAAA.....TATGATACGAGCT 1304





315 AlaValThrAlaThrAlaProLysAlaGlyProAspValLysProAlaVal 331  
500 .....  
331 lAlaValCysAlaGluAlaLysProAlaProProProProGlnGlnL 348  
501 TGCCCAAAATATCCGCTCAACCTGACCGACA ..... 532  
348 euProLysAlaAlaAlaAlaAlaProThrGlyThrGluLeuLysPro 364  
533 ACCGCAGACCGGACACGGCTTCCG ..... ACCCTTTC 567  
365 AlaThrAlaProProHisGlySerProArgAlaAsnSerHisThrValTh 381  
568 CACAATGCCGGCTATGCTGACGCAAGGAGTAGGCGACG ..... 607  
381 rValThrProProAsnValProArgAlaAlaAlaAlaThrValProThrA 398  
608 ..... GATTCAAACGGCCACCGATACAGCCCG 637  
398 lAglyAlaValProLysAlaSerThrGlyThrThrProAlaAlaAlaPro 414  
638 AGCTGG...ACAGATCGGGCAATCGCGCAAGCCTTCAACGGCAGCTGCA 684  
415 GlnGlnProValProLysAlaAlaProValThrProSerProGlnGln 431  
685 GATATCGTCAAAACATCATCGCGCGCGCAGGAGAAATTCGCGCGCAGG 734  
431 nAlaValProArgAlaAlaThrAla ..... AlalaAlaProv 444  
735 CGATCGCTGCAGGTATPAACGAAGCTCAACATTCGTGTCATGCACG 784  
444 alThrProGlnGlnProValThrLysAlaAlaLathThrThrAsnAlaThr 460  
785 GCTTGGGCTGCTTCCACCGAAACAGATGGCGCGCATCAACGATTTG 834  
461 Pro ..... ProGlnProLleProLysAlaAlaThrThr 473  
835 GCAGATATGGCGCACTCAAGACATATCGCGCAGCAG ..... 871  
473 rThrAla ..... ThrProValThrProGlnGlnProIleProLysA 487  
871 ..... 871  
487 lAglyThrAspAlaAlaProProProAlaValProLysAlaProSerAsp 503  
872 ..... CCATCCGCGATTGGGCGAGTCCAAAC 897  
504 GlyArgAlaAlaThrProGlyValProAsnAlaLathThrAspProGlnLy 520  
898 CCCAATCGCGCACGAAGCATAGAAGCGGTGAGCAATATCTTTATGGCAG 947  
520 sProProProThrProGlnSerValProSerAlaValThrGluProLysP 537  
948 CATCCCATCAAGAGGATTGGAGCTGTCGCGGGGAAATACGCTTGGCG 997  
537 rGlnPro ..... 539  
998 GCATCAGCGCACATCTCTCAAGCGGTGCGAGATGGCGGATCGCATTTG 1047  
539 ..... 539  
1048 CCGAAAGGAATCCGCGGTGAGCACAATTTGCCGATCGCGCATACGC 1097  
540 rArgAlaAlaProProProSerAsnGluAlaThrProAlaValProSerP 556  
1098 CAATACCGGT ..... CCCCTT...ACCATTCCCGAA ..... 1126  
556 rOserProAsnLeuLysSerProLeuProThrIleProLysProValPro 572  
1127 ..... ATATCGGTTCAACATTTGGAGCAGCGT 1152  
573 LeuMetAlaLeuThrProGlnProValThrAlaGlnMetValThrGlnLe 589

```

1153  TACGGCAAGAAACAATCACCTCTCAACCGTCGCCCGCTCAACAGCGCAA 1202
      ::::: ||| ||||| |||
589  uAlaAaThrLysProSerProIleValProLysAlaSerProLysAlaL 606
      ::::: ||| ||||| |||
1203  AAATGTCA 1210
      ||:::
606  euMetThr 608
      ||:::

seq_name: pir2.H87520

seq_documentation_block:
  Hypothetical protein CC2193 [imported] - Caulobacter crescentus
  C:Species: Caulobacter crescentus
  C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change
  C:Accession: H87520
  R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; E-
    B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn,
    n, J.; Ermolova, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Ve-
    proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
  A:Title: Complete Genome Sequence of Caulobacter crescentus.
  A:Reference number: A87249; MUID:21173698; PMID:11259647
  A:Accession: H87520
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-424 <STO>
  A:Cross-references: GB:AE005673; NTD:gl3423694; PIDN:AAK24164.1;
  C:Genetics:
  A:Gene: CC2193

```

```

alignment_scores:
  Quality: 133.00      Length: 489
  Ratio: 0.627        Gaps: 23
  Percent Similarity: 43.354  Percent Identity: 21.677

alignment_block:
US-09-303-518D-463 x H87520 ..

Align seg 1/1 to: H87520 from: 1 to: 424

303 CCATTGCGCCTTCGACAACCATGCTCACATTCCGATTCGTGACGAAGCG 352
|||||:|||||:|||||:
5 ProPheValSerArgArgPro..... 11

353 GTACTCCCGTTGACGGATTGAGCCTTTACCGCATCCATTGGGACGGATAC 402
|||||:|||||:|||||:
12 ..... 15 ArgAlaProAlaArgAlaArgAlaArgAlaArgProGlyProPheAlaLeu 31
|||||:|||||:|||||:
444 ..... CGGTATCCCGTCCCAAGCGCGGAGGA.. 473
|||||:|||||:|||||:
32 SerValAlaAspArgValArgCysValArgProGlnGlyArgHisGlyAl 48
|||||:|||||:|||||:
474 .....TATATACAGCTACGACATNAAG.....CGTTGCC 504
|||||:|||||:|||||:
48 acGlyLeuHisArgGluLeuLeuHisAlaArgGlyGlyAspGluIleA 65
|||||:|||||:|||||:
505 CAAATATCCGCTCACTGACCGGACACCGGACACCGGACACCGGCT 554
|||||:|||||:|||||:
65 rgArgLeuArgArgAspArgAlaArgTrpArgProHisArgProGlySer 81
|||||:|||||:|||||:
555 TGCGGACCGTTTCCACAATCGCGCTATGCTGACGCAAGGAGTAGGCG 604
|||||:|||||:|||||:
82 GlyLeuHisAlaArgAlaCysAlaGlyGlyGlyAspArgGluAlaAr 98
|||||:|||||:|||||:
605 ACGGATTCAACGGCCAC.....CGATACAGCCCCGAGCT...GGAC 645
|||||:|||||:|||||:
98 qHisLeuSerAspHisHisArgSerAlaArgArgProArgGlyGlnGly 115
|||||:|||||:|||||:

```



```
646 AGATCGGGCAATGCGCGGCAAGCCCTTCAACGCCACTCGCAGATATCGTCAA 695
      ::::: ||| ::::: |||:::
115 LylGlyGlnGlyArgValGlyLeuArgGlnGlnGlyAlaGly 131
      ::::: ||| ::::: |||:::
696 AAACATCATCGCGCGGCGGAGGAAATGTGCGCGCAGCGATGCCGTCG 745
      ::::: ||| ::::: |||:::
132 ArgArgHisArgHisAlaGluProAspProGlyArgGlyLeuCysAspAl 148
      ::::: ||| ::::: |||:::
746 AGGTATAGCAAGCGCTCAACATTTGCTGTCATCGACGGTGTGGTCTG 795
      ::::: ||| ::::: |||:::
148 a.....ArgArgAlaGlnArgCys.....GlyLeuGlyAlaA 159
      ::::: ||| ::::: |||:::
796 CTTTCACGAAACAGATGCGCGCATCAACGATTTGGCAGATATGCG 845
      ::::: ||| ::::: |||:::
159 la.....ArgGlnGlySerGlyValPro.....GlyArgAspPro 170
      ::::: ||| ::::: |||:::
846 GCAACTCAAGACTATGCGCGAGCAGCATCGCGATTTGGCAGTCCAAA 895
      ::::: ||| ::::: |||:::
171 SerGlnGlnAspArgHisArgProGlyArgArgValArgGlyAlaValAr 187
      ::::: ||| ::::: |||:::
896 ACCCAATCGCGCACAGGCATAGAGCGCTCAGCAATATCTTTATGCA 945
      ::::: ||| ::::: |||:::
187 gArgGlnAlaArgGlyArgPro.....AlaA 196
      ::::: ||| ::::: |||:::
946 GCATCCCCCATCAAGGATTTGGAGCTGTCCGGGGAATACGGCTTGG 995
      ::::: ||| ::::: |||:::
196 snHisProAlaArg.....ArgArgProLeuArg 206
      ::::: ||| ::::: |||:::
996 CGGCATCGGCAGCATCTGTCACGGTTCGAGATGGCGGATCGCAT 1045
      ::::: ||| ::::: |||:::
207 ArgHisGlnGlyGlyAlaGlnAspAla..... 215
      ::::: ||| ::::: |||:::
1046 TGCCGAAGGGAATCCCGCGTACGACAAATTTCCGATCGCGC..... 1091
      ::::: ||| ::::: |||:::
216 .....GluGluArgArgGlnArgProLeuArgArgProGlyArgA 229
      ::::: ||| ::::: |||:::
1092 ..ATAGCCAAATACCGCTCCCTTACCATTCCCGAAA.....TAT 1130
      ::::: ||| ::::: |||:::
229 laValGlyArgLeuProAspProValAlaLeuProArgArgAlaArg 245
      ::::: ||| ::::: |||:::
1131 CGTTCAACTTGGAGCGCTTACGGCAAGAAACATACCTCTCAA 1180
      ::::: ||| ::::: |||:::
246 ProArgProAlaGlyProGluLeuHisLeuArgProAsp.....LeuAr 260
      ::::: ||| ::::: |||:::
1181 CGGTGCGCGCG..... 1191
      ::::: ||| ::::: |||:::
260 g..ileProLeuSerGluArgProValMetLysThrArgThrLeuLeu 276
      ::::: ||| ::::: |||:::
1192 .....TC 1193
      ::::: ||| ::::: |||:::
277 ThrAlaAlaValAlaLeuSerLeuAlaGlyAlaAlaHisAlaAlaSe 293
      ::::: ||| ::::: |||:::
1194 AAACGGCAAAATGTCAACTGGCAGACAGCCAGCCAGCAGAGAGCGG 1243
      ::::: ||| ::::: |||:::
293 rHisAlaArgAspIlePheIleLysGluGlnAspGln..... 305
      ::::: ||| ::::: |||:::
1244 TACCGTTTCACGTAAGGTTTCCGAATTTTGAGAAGCAC..... 1284
      ::::: ||| ::::: |||:::
306 .....AsnGlyAspGlyPheValThrLysAspGlyrAlaAlaThr 319
      ::::: ||| ::::: |||:::
1285 .....GTGAATATATACGAAAGCTCGATATTCAGAAAT 1319
      ::::: ||| ::::: |||:::
320 ArgAlaIleGlnPheAlaLysThrAspThrAsp..... 330
      ::::: ||| ::::: |||:::
1320 ATCGGGGCGGTATACCTAAGCTTAAAGCTGTGTTGATGCGAAACCGA 1369
      ::::: ||| ::::: |||:::
331 .AsnSerGlyAlaLeuSerGlnAlaGlyrValAlaGluPheLysAlaA 347
      ::::: ||| ::::: |||:::
1370 GATGGGAGGTGATAGGAAGCTTAATAATTGACAACCTCGTGACGAGTG 1419
      ::::: ||| ::::: |||:::
347 rgLeuGluAlaLysLeuThrAlaSerAspMetThrProGluLysLysAla 363
      ::::: ||| ::::: |||:::
1420 GAGAAAATGTTTCAGCAACGAGAGAGG..... 1449
```

```
||||:||||:||||:||||:
364 GluGluArgMetArgGlnMetArgGlnThrAspValArgPheGlyValle 380
      |||:||||:||||:||||:
1450 ....AGTCAGAGTAGTCAGTTTAAAGCCCATGCGCACGAGATGGGAAA 1495
      |||:||||:||||:||||:
380 uAspSerAspLysSerGlyGlyIleThrLysAlaGluPheAspTyrSerG 397
      |||:||||:||||:||||:
1496 ATAAACAGGGTTAGATTTTAAATCATTTATAGTGGTGGTATCATCAATAAG 1545
      ::::: ||| ::::: |||:::
397 ly.....TrpArgMetPheValThrHisAspThrAsnThr 408
      ::::: ||| ::::: |||:::
1546 AAAGGCACAGTAACA 1560
      |||:||||:
409 AspGlyAlaValSer 413

seq_name: pir2:A29130

seq_documentation_block:
beta-amylase (EC 3.2.1.2) / alpha-amylase (EC 3.2.1.1) precursor - Bacillus polymyxa
C:Species: Bacillus polymyxa
C:Date: 25-Oct-1987 #sequence_revision 03-Mar-1994 #text_change 15-Oct-1999
C:Accession: A29130; B29130; A32251; A29108
C:Rawazu, T.; Nakanishi, Y.; Uozumi, N.; Sasaki, T.; Yamagata, H.; Tsukagoshi, N.; Ud
J. Bacteriol. 169, 1564-1570, 1987
A:Title: Cloning and nucleotide sequence of the gene coding for enzymatically active
A:Reference number: A29130; MUID:87165765
A:Accession: A29130
A:Molecule type: DNA
A:Residues: 1-936 <RAW>
A:Cross-references: GB:M15817
A:Experimental source: strain 72
A:Accession: B29130
A:Molecule type: protein
A:Residues: 36-50 <RA2>
R:Dozumi, N.; Sakurai, K.; Sasaki, T.; Takekawa, S.; Yamagata, H.; Tsukagoshi, N.; Ud
J. Bacteriol. 171, 375-382, 1989
A:Title: A single gene directs synthesis of a precursor protein with beta- and alpha-
A:Reference number: A32251; MUID:89123046
A:Accession: A32251
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 937-1196 <UOZ>
A:Experimental source: strain 72
R:Rhodes, C.; Strasser, J.; Friedberg, F.
Nucleic Acids Res. 15, 3934, 1987
A:Title: Sequence of an active fragment of B. polymyxa beta amylase.
A:Reference number: A29108; MUID:87231094
A:Accession: A29108
A:Molecule type: DNA
A:Residues: 'MIGL', 2-66, 'S', 68-99, 'D', 101-153, 'N', 155-176, 'Q', 178-226, 'KS', 229-329, 'S
'X', 737-740, 'S', 742-757, 'N', 759-776 <RHO>
A:Experimental source: ATCC 8523
C:Genetics:
A:Start codon: TTG
C:Function:
A:Description: catalyzes both the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
vely from the non-reducing end (beta-amylase activity)
A:Pathway: glycogen/starch degradation
C:Superfamily: alpha-amylase core homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-35/Domain: signal sequence #status predicted <SIG>
F;36-1196/Product: beta-amylase/alpha-amylase #status predicted <MAT>
F;908-1033/Domain: alpha-amylase core homology <AMY>
```

```
alignment_scores:
Quality: 133.00 Length: 717
Ratio: 0.426 Gaps: 38
Percent Similarity: 43.515 Percent Identity: 21.060

alignment_block:
US-09-303-518D-463 x A29130 ..
```

Align seg 1/1 to: A29130 from: 1 to: 1196

```
71 CCTCAGATTGGCAACAGCATCCTTTATCCGGCAGGTTCTCGACGGTCAG 120
|||||:|||||:
187 ProLysileTyr.....LeuSerGly..... 193
121 CATTTCCGACCCGACCGGGAAT.....ACCACCTATTCCGGCAG 158
|||||:|||||:
194 .....GlyProSerGlyGluLeuArgTyrProSerTyrProAlaAa 208
||||| ||| |||:
159 CAGGGGGGAGCTTCCGNAGC..... 178
||||| ||| |||:
208 laGlyTrpSerTyrProGlyArgGlyLysPheGlnAlaTyrThrGluThr 224
179 .....GCAACGGCCATATCGGAT..... 196
|||||:|||||:
225 AlaLysAsnAlaPheArgThrAlaMetAsnAspLysTyrGlySerLeuAs 241
197 .....TGGAAACATACAAAGCCATCA.....GTT 221
||||| ||| |||
241 pLysIleAsnAlaAlaTrpGlyThr.LysLeuThrSerLeuSerGlnIle 257
222 GGCCACCTCGATGAT.....TCAACAGGGCGCGCTTGAAG 256
||||| ||| |||
258 AsnProThrAspGlyAspGlyPheTyrThrAsnGlyGlyTyrAsnSe 274
||||| ||| |||
257 GAAATATCGCTCATPTGTCGCTTTCGATCACGGGCACAAATCCAT 306
||||| ||| |||
274 rAlaTyr.Gly.....LysAspPheLeuSerTyrTrpGln 285
307 TCGCCCTTCGACACCAT.....GCCTCATCTCGA 338
|||||:|||||:
286 SerValLeuGluLysHisLeuGlyValIleGlyAlaAlaHisLysAs 302
339 TTCCTGACGAAGCGGTAGTCCGCTTACGACGATTCAGCCCTTACCGCATCC 388
|||||:|||||:
302 nPheAspSerValPheGlyValArgIleGlyAlaLysIleSerGlyLeuH 319
389 ATTGGACGGATACGAACACCATCCCGCGGCGGTATGACGGGCACAG 438
|||||:|||||:
319 istrpGln.....MetAsnAsnProAlaMetProHisGlyThrGluGln 333
439 GCGGGGGCTATCCCGCTCCCAAGCGGAGGATATATACAGCTACGA 488
|||||:|||||:
334 AlaGlyGlyTyr.....TyrAspIleAs 341
489 CATAAAGCGGTTCGCCAAATATCCGCTCAACCTGACCGACACCGCA 538
341 n..... 341
539 GCACCGGACACGGGTGTCGACCGTTTCCACATGTCGCGGCTATGCTG 588
|||||:|||||:
342 .....ArgLeuIleGlnLysPheLysAspAlaAspLeu 354
589 ACCGAGGAGTAGCGCACGGATTCACCGCCCGCCCGATACAGCCCGCA 638
||||| ||| |||
355 .....ThrPheThrCysLeuGln 360
639 GCTGCACAGATCGGCAATGCGCGGAGCCCTC.....AACGGCAGCTG 682
|||||:|||||:
360 uMetSerAspSerGlyThrAlaProAsnTyrSerLeuProSerThrLeuV 377
683 CAGATATCGTCAAAACATCATCCGCGCGGCAGGA..... 717
|||||:|||||:
377 aLAspThrValSerSerIleAlaAsnAlaLysGlyValArgLeuAsnGly 393
718 .....GAAATTCGCGCGCAGCATGCCGTGCGAGGTATAGCGAAGG 761
|||||:|||||:
394 GluAsnAlaLeuProThrGlyGlySerGlyPheGlnLysIleGluGlu 410
762 CTCAACATTTGCTGTCATCGCGGTGCTGCTCTTCCACCGCAACA 811
|||||:|||||:
410 sIleThrLysPheGlyTyrHisGlyPheThrLeuLeu..... 422
```





```

544  LeuProProHisSerGlnValSerTyrSerGlnAlaGlyProAsnG1 560
      |||::: ||||| |||
1044 ATTGCCGAAGGAAATCCGCC...GTCAAGCAATTTGCCGATCCGG 1090
      ||| :::::|||||
560  yProProValSerSerSerAsnSerSerSerThrSerGlnGlyS 577
      |||::: |||||
1091 CATAC.....GCAAA 1101
      ::||| |||
577  erTyrProCysSerHisProSerProSerGlnGlyAlaPro 593
      |||::: |||||
1102 TACCGTCCCTTACCATTCCGAAATA.....TCGTTCAACATTGGA 1145
      ||||| ||| ::||| ::|
594  TyrProPhePro.ProValProThrValThrSerSerAlaThrLeuS 610
      |||::: |||||
1146 GCAGCGTTACGGCA.....AAGAAACA 1168
      |||::: |||||
610  erThrValIleAlaThrValAlaSerSerProAlaGlyTyrLysThrAla 626
      |||::: |||||
1169 TCACCTCGCTCAACCG 1183
      |||::: |||||
627  SerProGlyPro 631
      |||::: |||||

```

seq\_documentation\_block:  
 hypothetical protein Vng1794c [imported] - Halobacterium sp. NRC-1  
 C.Species: Halobacterium sp. NRC-1  
 C.Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C.Accession: B84331  
 R.Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.  
 ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A: 97, 12176-12181, 2000  
 A.Proc: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
 A.Title: Genome sequence of Halobacterium species NRC-1.  
 A.Reference number: A84160; MUID:20504483  
 A.Accession: B84331  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-694 <STO>  
 A.Cross-references: GB:AE004437; NID:g10581248; PIDN:AAG20014.1; GSPDB:GN00138  
 C.Genetics:  
 A:Gene: VNG1794C

alignment_scores:		
Quality:	131.50	Length: 410
Ratio:	0.739	Gaps: 23
Percent Similarity:	43.415	Percent Identity: 24.146

alignment\_block:

US-09-303-518D-463 x B84331

Align seg 1/1 to: B84331 from: 1 to: 694

116 GTCAGCATTTCCGAACCGCAGGGAAATACCACCTATTTCGCCAGCAGGCGG 165  
|||:: : :::::|||::: ||||| ::: |||::  
72 ValAlaAlaGluProThrAla...ThrThrValAspArgAspGlyAs 87  
::: : :::::|||::: ||||| ::: |||::  
166 GAGCTGCCNAGCGCACGCGCCATATCGGATTCGGAAACAATACAAA.... 211  
::: : :::::|||::: ||||| ::: |||::  
87 nAsp.....AspThrThrAlaTyProAspGlyGluThrIleProLeuM 102  
212 .....GCCTACGTTCGGCGCACTCGA 232  
|||:: : :::::|||::: ||||| ::: |||::  
102 etAlaValAspGlySerValValAlaPheAlaLaProPheAla.ProAs 118  
233 TGATTTCAACAGGCGCGTTCAGGAATAATCGCTACATTGTCGC... 279  
|||||:: : ::|||:: ||||| ::: |||::  
118 pAspSerAspAlaGlySer..TyrGlyAsnGluAspValLeuLeuAsnVal 134  
280 .....TTTTCCGATCAGGGCACAAATC.....CATTC 308

